NOVEL FRONTIERS IN HELMINTH GENOMICS

EDITED BY: Jose F. Tort, Gabriel Rinaldi, Makedonka Mitreva and
Klaus Rüdiger Brehm
PUBLISHED IN: Frontiers in Genetics

frontiers Research Topics

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

NOVEL FRONTIERS IN HELMINTH GENOMICS

Topic Editors:
Jose F. Tort, Universidad de la República, Uruguay
Gabriel Rinaldi, Wellcome Sanger Institute (WT), United Kingdom
Makedonka Mitreva, Washington University School of Medicine in St. Louis, United States
Klaus Rüdiger Brehm, Julius Maximilian University of Würzburg, Germany

Table of Contents

frontiers
in Genetics

Check for
updates

Editorial: Novel Frontiers in Helminth Genomics

Jose F. Tort¹*, Makedonka Mitreva², Klaus R. Brehm³ and Gabriel Rinaldi⁴

¹ *Department of Genetics, Faculty of Medicine, University of the Republic, Montevideo, Uruguay, ² Department of Medicine, Washington University School of Medicine, St. Louis, MO, United States, ³ Institut für Hygiene und Mikrobiologie, Julius Maximilian University of Würzburg, Würzburg, Germany, ⁴ Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, United Kingdom*

Editorial on the Research Topic

Novel Frontiers in Helminth Genomics

At the brink of the third decade of the twenty-first century, almost a quarter of the world's population are infected by parasitic worms, mainly in deprived areas worldwide. Despite huge and promising efforts from the World Health Organization, there is still much road to cover for the control of these highly prevalent neglected diseases. The current issue highlights how the knowledge of helminth genomics is being translated into novel strategies for diagnosis, control, and the general understanding of parasite biology and adaptation.

Determining infection burdens precisely is critical for surveillance, to evaluate the impact of control programs and to make informed decisions in helminth eradication programs. Two articles in this issue demonstrate how genomics can contribute to these efforts. Grant et al. show that it is feasible to exploit available genomic information to identify species-specific repeated sequences in order to generate qPCR-based diagnostic assays for diverse nematodes. A comparison between these approaches and traditional parasitological methods in soil-transmitted helminths field tests is evaluated and discussed.

Hedtke et al. highlight the relevance of the epidemiological perspective based on genomics in control program decision making. For filarial parasites in Africa, they show that the combination of genomic epidemiology and genome-wide associations could be helpful in defining transmission zones, identifying local or introduced parasites as sources of reinfection and distinguishing genetic markers associated with parasite response to chemotherapy. These are key elements to elaborate surveillance and eradication strategies.

A relevant aspect for these epidemiological approaches is the ability to obtain relevant parasitic samples from natural infections, usually in difficult field conditions where the quantity and quality of samples are limited. Doyle et al. describe the optimization of different methods for extraction and sequencing of high-quality DNA from minimal amounts of parasitic material in several nematodes and trematodes. Although difficult, the authors prove that it is feasible to extract DNA from single parasite eggs and early larval stages, a breakthrough in epidemiological surveillance and genetic population studies, among other relevant applications.

Host-parasite interactions have always caught the attention of parasitologists. It is well-documented that parasites modulate the immune responses of their hosts, but in this issue Martin et al. extend this view by including the gut microbiome. Based on a randomized placebo-controlled anthelminthic trial, the work is the first to analyze changes in the gut microbiome, the presence of parasitic helminths and whole blood cytokine responses in parallel, and opens the avenue to a novel and more complex picture of host-parasite interaction.

The work of Jasmer et al. is a good example of multi-omics approaches applied to understand the complexity of parasitic helminths at the molecular level. They focus on the nematode intestine as

a target for therapies, collecting and combining genomics, transcriptomics, proteomics, and miRNA profiling data from representative species of different nematode clades that live and feed on diverse locations within the host digestive tract. The review provides a comprehensive view of the advancements, methods and resources available to analyze gene and protein expression and regulation on the nematode intestine.

Translational research that involves mining available genomic data to understand relevant biological processes is exemplified by the work of Otero et al.. They have taken advantage of available genomes to analyze the basis of alternative mitochondrial electron transport chains used by helminths in hypoxic conditions, showing that although the mechanism seems to be conserved, different gene duplications of the genes involved in the pathway are evident in diverse lineages of both nematodes and flatworms, and evaluated their significance with functional analysis in *C. elegans*.

Reliable methods of functional genomics are needed for analyzing gene function. Lok reviews the methodologies in DNA and RNA transformation in parasitic nematodes, and the seminal works that set proof of principles of CRISPR-CAS9 mutagenesis in strongilyd worms. The value of dominant mutations and regulable Cas9 expression in gene functional validation in parasitic nematodes are thoroughly discussed.

Two articles in this issue are focused on the use of genomics to analyze evolutive aspects of helminths. Compositional features of flatworm genomes are analyzed in detail by Lamolle et al.. Extreme biases in base composition are correlated to profound preferences in codon usages in diverse lineages, with differences between and within free living and parasitic flatworms. The evolutionary implications of these findings are discussed.

By sequencing the tapeworm *Echinococcus oligartus,* and comparing with other species responsible for cystic, multicystic and alveolar echinococcosis, Maldonado et al. provide a novel genomic perspective of cestode evolution, showing that this sylvatic species responsible for unicystic echinococcosis is basal to the genera, and highlighting the relevance of studying sylvatic species.

The use of transcriptomics approaches to shine a light on diverse biological phenomena in helminths is illustrated by works on trematodes. One of the more intriguing aspects in helminth biology is the existence of separate sexes in schistosomes, and the need of pairing for female sexual maturation. While several efforts have focused on the egg-producing female, Lu et al. review the male partner of the couple, highlighting genes related to nervous system and signaling pathways in the complex nature of this interaction.

Helminth parasite survival is supported by massive egg production. Grossi Gava et al. use a combination of RNAi and transcriptomics to demonstrate the involvement of cJun terminal kinases in egg production. Again, using comparative genomics they show that orthologous genes in the nematode *C. elegans* are related to sterility and oocyte maturation.

Regulation of gene expression by diverse RNA mediators is now a central topic in biology, and helminths are no exception. Maciel et al. investigated long non-coding RNAs (lncRNAs) in *S. mansoni* by reanalyzing more than 600 RNAseq libraries. Their extensive characterization shows that lncRNA expression is regulated epigenetically, and they identified sets of correlated gene co-expression modules comprising lncRNA and mRNAs across diverse stages.

In brief, this Research Topic highlights the diversity and richness of genomic-based perspectives used to advance the study of critical aspects of the biology of helminths, that still today inflict much suffering to the most deprived in the world.

AUTHOR CONTRIBUTIONS

CRISPR/Cas9 Mutagenesis and Expression of Dominant Mutant Transgenes as Functional Genomic Approaches in Parasitic Nematodes

James B. Lok*

Department of Pathobiology, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, PA, United States

DNA transformation of parasitic nematodes enables novel approaches to validating predictions from genomic and transcriptomic studies of these important pathogens. Notably, proof of principle for CRISPR/Cas9 mutagenesis has been achieved in *Strongyloides* spp., allowing identification of molecules essential to the functions of sensory neurons that mediate behaviors comprising host finding, invasion, and location of predilection sites by parasitic nematodes. Likewise, CRISPR/Cas9 knockout of the developmental regulatory transcription factor *Ss-daf-16* has validated its function in regulating morphogenesis of infective third-stage larvae in *Strongyloides stercoralis*. While encouraging, these studies underscore challenges that remain in achieving straightforward validation of essential intervention targets in parasitic nematodes. Chief among these is the likelihood that knockout of multifunctional regulators like *Ss*-DAF-16 or its downstream mediator, the nuclear receptor *Ss*-DAF-12, will produce phenotypes so complex as to defy interpretation and will render affected worms incapable of infecting their hosts, thus preventing establishment of stable mutant lines. Approaches to overcoming these impediments could involve refinements to current CRISPR/Cas9 methods in *Strongyloides* including regulatable Cas9 expression from integrated transgenes and CRISPR/Cas9 editing to ablate specific functional motifs in regulatory molecules without complete knockout. Another approach would express transgenes encoding regulatory molecules of interest with mutations designed to similarly ablate or degrade specific functional motifs such as the ligand binding domain of *Ss*-DAF-12 while preserving core functions such as DNA binding. Such mutant transgenes would be expected to exert a dominant interfering effect on their endogenous counterparts. Published reports validate the utility of such dominant-negative approaches in *Strongyloides*.

Keywords: transgenesis, parasitic nematode, CRISPR/Cas9, dominant transgene, mutagenesis

INTRODUCTION

The advent of transgenesis in parasitic nematodes has opened new means of ascertaining the functions of specific genes in these important pathogens by both gain- and loss-of-function approaches. The first successes in this area involved translating methods for transgenesis in *Caenorhabditis elegans* to parasites in the genera *Strongyloides* and *Parastrongyloides*, which were apt subjects given their

capacity to undertake one or more generations of free-living development. Transgenesis in *Strongyloides* spp. has enabled targeted mutagenesis by CRISPR/Cas9, and this method has already found application in studies that assigned function to genes regulating morphogenesis and host-finding behavior in infective larvae. Transgenes designed to exert dominant interfering effects over specific genes have also revealed essential functions of these genes in growth and developmental of pre-parasitic larvae of *Strongyloides stercoralis*. Robust new methods for integrative transgenesis in *Brugia malayi* promise to open these and other avenues of functional genomic study in the filariae and potentially in other obligately parasitic nematodes. None of these functional genomic approaches are new to the basic fields of cellular and molecular biology, but their application to parasitic helminths marks a recent innovation that ranks among the salient advances in parasitology in the past two decades. In view of this, the following review discusses these current findings in detail and proposes new directions in their application to increasingly informative and practical studies of gene function in a group of pathogens that impose the risk of morbidity and mortality on a large segment of the world's population and exact a heavy toll on the health and welfare of domestic animals.

TRANSFORMATION WITH DNA OR RNA CONSTRUCTS IS A PREREQUISITE TO DEPLOYING CONTEMPORARY METHODS FOR ASSESSMENT OF GENE FUNCTION IN PARASITIC NEMATODES

Approaches to Transgenesis in Clade III and IV Parasitic Nematodes
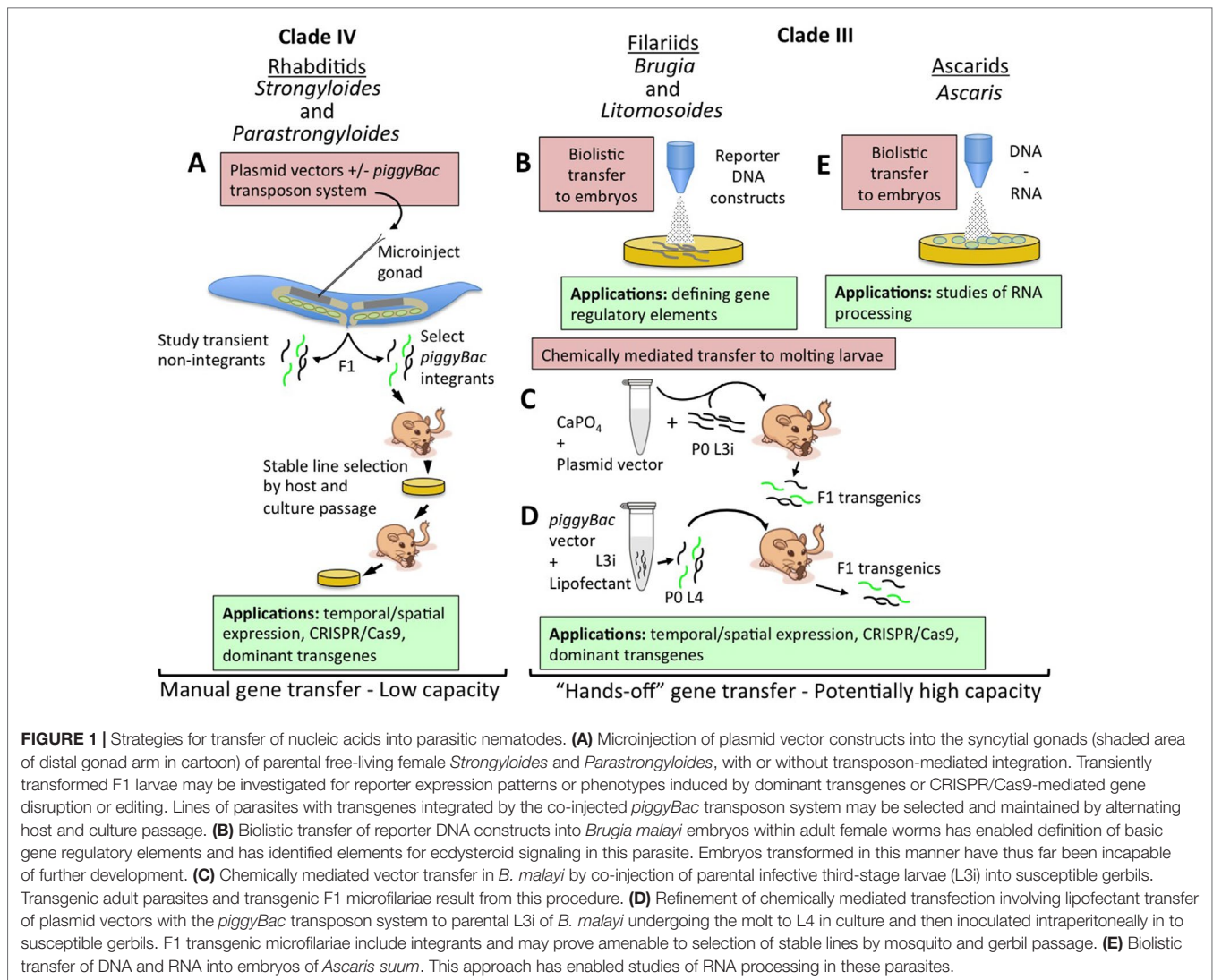
Parasites in the genus *Strongyloides* and *Parastrongyloides* were the first to be successfully transformed with plasmid-based constructs (Lok and Massey, 2002; Grant et al., 2006a; Li et al., 2006; Li et al., 2011). The main factor allowing this advancement was these parasites' unique ability to execute one or more generations of development as free-living males and females with their progeny in the soil (Viney and Lok, 2015; Grant et al., 2006b). These free-living cycles provide access to the adult germlines of *Strongyloides* and *Parastrongyloides* that is unavailable in obligately parasitic nematodes. Free-living generations of *Strongyloides* and *Parastrongyloides* can be maintained in agar plate culture using methods adapted from those used for the free-living nematode *C. elegans* (Grant et al., 2006a; Lok and Unnasch, 2013). *Parastrongyloides trichosuri* can execute an indefinite number of sequential free-living generations in plate culture, enhancing its utility as a subject for genetic study. Most crucially, the body plans of free-living female *Strongyloides* and *Parastrongyloides* and of *C. elegans* hermaphrodites are so similar that it was straightforward to adapt methods for gene delivery by **gonadal microinjection** from *C. elegans* (Fire, 1986; Fire et al., 1990a; Fire et al., 1990b; Mello et al., 1991; Mello and Fire, 1995) to the parasites (**Figure 1A**) (Lok and Massey, 2002; Grant et al., 2006a; Li et al., 2006; Li et al., 2011). A similar approach was used to achieve transgenesis in *Strongyloides stercoralis* by microinjection

of plasmid constructs into the testes of free-living males (Shao et al., 2017). These approaches, involving transfer of plasmid-based vector constructs, allow promoter-regulated, tissue-specific transgene expression in F1 transformants (**Figure 1A**) (Grant et al., 2006a; Li et al., 2006; Junio et al., 2008; Li et al., 2011; Shao et al., 2017), and integration of transgene sequences into the chromosomes of *Strongyloides ratti* permits establishment of stable transgenic lines of this parasite (**Figure 1A**) (Shao et al., 2012). These methods have enabled studies of gene function in *Strongyloides* and *Parastrongyloides* that have revealed anatomical and temporal patterns of gene expression and essential functions of specific genes in development and survival of the pre-infective stages of these parasites (**Figure 1A**) (Massey et al., 2003; Grant et al., 2006a; Junio et al., 2008; Castelletto et al., 2009; Stoltzfus et al., 2012; Massey et al., 2013; Yuan et al., 2014a; Yuan et al., 2014b; Lei et al., 2017; Yuan et al., 2017).

Significant strides have also been made towards transgenesis in the filaria *Brugia malayi*. This parasite, and all filariae, are far more challenging subjects for molecular genetic study, as they have no free-living generation comparable to *Strongyloides* and related genera, and furthermore, unlike soil transmitted parasitic nematodes, their pre-infective stages require an arthropod vector for development. These barriers notwithstanding, transient DNA transformation of *B. malayi* (Higazi et al., 2002; Higazi and Unnasch, 2013) and *Litomosoides carinii* (Jackstadt et al., 1999) has been achieved by particle bombardment or biolistics (**Figure 1B**). In the case of *B. malayi*, the primary object of this approach has been to investigate the structure and function of gene regulatory sequences by transforming embryos within the uteri of female worms with plasmid constructs encoding fluorescent reporters linked to wild-type or mutant regulatory sequences, promoters mainly, and evaluating function by quantifying reporter activation. These bombarded embryos are not developmentally competent, but this approach has, nevertheless, yielded a wealth of basic information on the structural motifs required for promoter activity (Shu et al., 2003; Higazi et al., 2005; de Oliveira et al., 2008; Bailey et al., 2011), trans-splicing (Higazi and Unnasch, 2004; Liu et al., 2007; Liu et al., 2010a), and operon function (Liu et al., 2010b) in *B. malayi*. This approach to transient transgenesis has also revealed the function of microRNAs in regulation of *B. malayi* genes (Liu et al., 2015), the presence of tetracycline regulatable promoters (Liu et al., 2011), and the transcriptional targets of ecdysone signaling in this parasite (Tzertzinis et al., 2010; Liu et al., 2012). Notably, the approach of transiently transfecting embryos has also been applied to *Ascaris suum* (**Figure 1E**) (Davis et al., 1999).

Stable, heritable transformation has been achieved in *B. malayi* by chemically mediated transfer of plasmid vector constructs into developing infective third-stage larvae (L3i; **Figures 1C, D**). Initial applications of this approach involved forming calcium phosphate co-precipitates with plasmid DNA constructs encoding a luciferase (GLuc) reporter (Xu et al., 2011). Mosquito-derived infective third-stage larvae (iL3) were then co-injected along with the calcium phosphate plasmid DNA co-precipitate into the peritoneal cavities of susceptible gerbils. Substantial proportions of *B. malayi* adults and their microfilarial progeny recovered from gerbils inoculated with transduced iL3

FIGURE 1 | Strategies for transfer of nucleic acids into parasitic nematodes. **(A)** Microinjection of plasmid vector constructs into the syncytial gonads (shaded area of distal gonad arm in cartoon) of parental free-living female *Strongyloides* and *Parastrongyloides*, with or without transposon-mediated integration. Transiently transformed F1 larvae may be investigated for reporter expression patterns or phenotypes induced by dominant transgenes or CRISPR/Cas9-mediated gene disruption or editing. Lines of parasites with transgenes integrated by the co-injected *piggyBac* transposon system may be selected and maintained by alternating host and culture passage. **(B)** Biolistic transfer of reporter DNA constructs into *Brugia malayi* embryos within adult female worms has enabled definition of basic gene regulatory elements and has identified elements for ecdysteroid signaling in this parasite. Embryos transformed in this manner have thus far been incapable of further development. **(C)** Chemically mediated vector transfer in *B. malayi* by co-injection of parental infective third-stage larvae (L3i) into susceptible gerbils. Transgenic adult parasites and transgenic F1 microfilariae result from this procedure. **(D)** Refinement of chemically mediated transfection involving lipofectant transfer of plasmid vectors with the *piggyBac* transposon system to parental L3i of *B. malayi* undergoing the molt to L4 in culture and then inoculated intraperitoneally in to susceptible gerbils. F1 transgenic microfilariae include integrants and may prove amenable to selection of stable lines by mosquito and gerbil passage. **(E)** Biolistic transfer of DNA and RNA into embryos of *Ascaris suum*. This approach has enabled studies of RNA processing in these parasites.

expressed GLuc, proving inheritance of the reporter transgene in the F1 generation (**Figure 1C**). This method was subsequently refined such that DNA transfer was achieved by lipofection of mosquito-derived L3i of *B. malayi* that were undergoing the L3i-L4 molt in co-culture with bovine embryo skeletal muscle cells (Liu et al., 2018). Such cultured larvae co-transformed with GLuc and reporters encoding fluorescent proteins and flanked by inverted tandem repeats specific for the *piggyBac* transposon system as well as a plasmid encoding the *piggyBac* transpose were capable of establishing patent infection in susceptible gerbils (**Figure 1D**). Analysis adult *B. malayi* and F1 microfilariae revealed transgene expression and also integration of transgene sequences into the parasite genome. Altogether, this represents a remarkable advancement in functional genomic methodology for filariae and for all parasitic nematodes. For the former, it marks the first integrative transformation of *B. malayi*, the most widely used model in experimental filariasis, and the first deployment of fluorescent reporters that will allow rapid visual selection of transgenic parasites and assessment of temporal and anatomical patterns of transgene expression. For all parasitic nematodes, the work on *B. malayi* identifies chemically mediated DNA transfer as an alternative to gonadal microinjection as a means of transgenesis. This method has the advantage of DNA transfer into the larval germline, making it applicable to obligate parasitic nematodes in which the adult germline is not as easily accessible as it is in the free-living adults of *Strongyloides* and related parasites. **Chemically mediated gene transfer** has the additional and paramount advantage of being a "hands-off" method not requiring the meticulous and painstaking process of microinjection and having the potential for high capacity transformation of hundreds or thousands of parasites in a single replication of the procedure (**Figure 1B–E**). The common factor in success of the two chemically mediated approaches used to date in *B. malayi* (Xu et al., 2011; Liu et al., 2018) appears to have been that DNA transfer was undertaken during an intermolt period, in this case the L3–L4 molt, in which the nascent cuticle was transiently competent to take up either the calcium phosphate–DNA co-precipitate or the

lipofectamine–DNA micelles. Parasitic nematodes such as hookworms or trichostrongyles and related rodent parasites such as *Nippostrongylus brasiliensis* and *Heligmosomoides polygyrus*, whose pre-parasitic larvae develop in the soil or on herbage, would seem to be particularly amenable to this approach, as they can be cultured through two molt cycles (L1–L3) in simple laboratory media. The same is true of *Strongyloides* and related genera, where the combination of life cycles containing both free-living and parasitic generations and a hands-off, high-capacity method of DNA transfer would make a particularly powerful model for molecular genetics in parasitic nematodes.

FUTURE DIRECTIONS IN TRANSGENESIS FOR PARASITIC NEMATODES

Gene Editing *via* Transgenesis in Pre-Parasitic Larvae of Soil-Transmitted Parasitic Nematodes

The microinjection-based methods of gene transfer into free-living females of *Strongyloides* spp. and of *Parastrongyloides trichosuri* have not been adapted to obligately parasitic nematodes whose pre-parasitic larvae develop in and are transmitted from the soil or vegetation and whose adult stages reside solely in vertebrate animal host. These include the hookworms and trichostrongyles of paramount importance to human medicine and animal agriculture. While the adult germlines of these parasites are not readily accessible for gene transfer, the rudimentary gonads of their pre-parasitic larvae, which are easily cultured in the laboratory, are accessible. Certainly, the chemically mediated and lipofectant-facilitated systems of gene transfer that have been so successful for *ex vivo* iL3 of *Brugia malayi* (Xu et al., 2011; Liu et al., 2018) should be investigated as possible approaches to gene transfer into hookworms and trichostrongyles of humans and domestic animals and their rodent counterparts that are important subjects for immunological research (Bouchery et al., 2017).

In addition to these "hands-off" approaches to gene transfer, it is also possible to transfer CRISPR/Cas9 elements into pre-parasitic larvae of soil-transmitted parasites by microinjection (Adams et al., 2019). Mutations were targeted to the *rol-6* ortholog in *S. stercoralis* by microinjecting a nucleoprotein comprising Cas9 and a target specific gRNA along with lipofectamine into the body cavities of iL3. This approach produced P0 mutant larvae exhibiting a classic "roller" phenotype identical to that seen in *rol-6* mutants of *C. elegans*. This approach might be adapted to the pre-infective larvae of other soil-transmitted parasitic nematodes as well.

Deployment of "Hands-Off", High-Capacity Gene Transfer in *Strongyloides* and Other Soil-Transmitted Parasitic Nematodes

A major shortcoming of gene transfer by gonadal microinjection in *Strongyloides* and related parasites has been the limited numbers of F1 transgenic larvae that can be produced by this approach

(**Figure 1A**). This creates difficulties in amassing samples of transgenic larvae of sufficient size to achieve statistical power and is a particular hurdle to the establishment of transgenic lines of *S. stercoralis*, where the most susceptible rodent host, the gerbil, requires an approximate minimum of 100 transgenic larvae to mount a patent infection yielding transgenic F2 progeny. This situation calls for a concerted effort to achieve chemically or lipofectant-mediated gene transfer into pre-infective larvae of *Strongyloides* spp. in a manner similar to the successful methods now available for *ex vivo* iL3 of *Brugia* spp. (Xu et al., 2011; Liu et al., 2018). It seems likely that these "hands-off" methods (**Figure 1B–E**) could be similarly deployed for gene transfer into pre-parasitic larvae of hookworms, trichostrongyles, and their rodent-infecting counterparts.

CRISPR/CAS9 MUTAGENESIS IS FEASIBLE IN PARASITIC NEMATODES

DNA transformation of parasitic nematodes opens the formal possibility of deploying CRISPR/Cas9 for targeted mutagenesis as a means of directly assessing gene function in these pathogens. Proof of principle for **CRISPR/Cas9 mutagenesis** has been achieved in *Strongyloides stercoralis* and *Strongyloides ratti*. This includes preliminary evidence of precisely targeted double-stranded DNA breaks (DSB) and homology-directed repair with a co-transfected DNA template (Lok et al., 2017) and comprehensive demonstration and phenotypic evaluation of targeted disruption of *Ss-unc-22* and *Sr-unc-22* in *S. stercoralis* and *S. ratti,* respectively (Gang et al., 2017). Mutations have been created in *Strongyloides* by transducing free-living females with basic CRISPR elements, including gRNA, Cas9 codon-optimized for expression in *Strongyloides*, and selectable markers encoded in plasmid vectors (Gang et al., 2017; Lok et al., 2017) or by microinjecting gonads of free-living females with pre-formed nucleoproteins comprising recombinant Cas9 and *in vitro* transcribed gRNAs (Gang et al., 2017; Adams et al., 2019). To date, transduction of parental worms with plasmid-encoded CRISPR/Cas9 elements has resulted in more efficient mutagenesis than is seen with delivery of pre-formed nucleoproteins (Gang et al., 2017).

Informative phenotypes have been associated with both large deletions at those sites mediated by non-homologous end joining (NHEJ) and with insertional mutagenesis *via* homology-directed repair (HDR) of CRISPR/Cas9-induced DSB. Most notably, a prominent "twitcher" phenotype, which is exacerbated by nicotine exposure, results from CRISPR/Cas9 disruption of *Ss-unc-22* and *Sr-unc-22*. These genes, like their *C. elegans* ortholog, encode the intracellular muscle protein twitchin (Moerman and Baillie, 1979; Moerman et al., 1988; Gang et al., 2017). The *unc-22* phenotypes in *Strongyloides* spp. are highly reminiscent of the twitcher phenotype resulting from mutations in C. elegans unc-22. Notably, the *C. elegans* twitcher phenotype is dominant, occurring in both homozygous and heterozygous *unc-22* mutants. Homology-directed repair has been exploited to integrate short disrupting oligonucleotides containing multiple

stop codons and novel priming sites for genotyping at the site of CRISPR/Cas9-induced DSBs (Lok et al., 2017) as well as a full reporter cassette encoding a fluorescent reporter that serves to both disrupt the target gene and to provide a convenient marker for mutant selection (Gang et al., 2017). In the latter instance, F1 larvae expressing the red fluorescent protein mRFPmars under the body wall-specific promoter for *Ss-act-2* from the repair template in an episomal array or from the template integrated into a CRISPR/Cas9 cut site in *Ss-unc-22* could be distinguished by PCR using primers specific for the repair template and the genomic region flanking the cut site and by the consistent pattern of body wall expression in integrants (Gang et al., 2017). Integrants in this case exhibited a uniform twitcher phenotype, providing proof of principle for the use of reporter-encoding repair templates to disrupt target genes by CRISPR/Cas9- and HDR-mediated integration.

Most significantly, CRISPR/Cas9-induced mutations in *S. stercoralis* are heritable and may be propagated by passage in gerbils, which constitute the most authentic rodent model of the entire spectrum of *S. stercoralis* infection and disease in humans (Nolan et al., 1993; Gang et al., 2017).

In addition to the studies of *Ss-unc-22*, which demonstrated that its encoded muscle component is required for normal motility of iL3 (Gang et al., 2017), CRISPR/Cas9 and the *Strongyloides* model have found application in studies of genes involved in the sensory biology of these infective larvae. Soil transmitted parasitic nematodes such as hookworms and *Strongyloides* spp. invade the host by skin penetration and develop as free-living pre-parasitic larvae in contaminated soil. iL3 of these parasites orient toward potential hosts using a complex of chemical and physical cues (Bryant and Hallem, 2018b; Castelletto et al., 2014). Orienting towards a thermal source is a behavior that many species of environmentally motile parasitic nematode iL3 use to acquire a host (Castelletto et al., 2014; Bryant and Hallem, 2018a; Bryant and Hallem, 2018b; Bryant et al., 2018). In *C. elegans*, thermal signals are received by receptor guanylate cyclases in sensory neurons of the amphidial complex, which are in contact with the external environment. These receptors signal through a heteromeric cGMP-activated cation channel, comprising TAX-2 and TAX-4 subunit proteins, that regulates polarization of the sensory neuron and potentiation of thermotactic responses (Komatsu et al., 1996; Bargmann and Mori, 1997; Komatsu et al., 1999). The *tax-4* ortholog in *S. stercoralis* has been disrupted by CRISPR/Cas9, and knockout iL3 exhibit diminished themotaxis towards 34°C compared to controls derived from parental worms receiving all CRISPR/Cas9 elements except the Cas9 endonuclease. This provides direct evidence that *tax-4* is required for normal thermotaxis by iL3 of this parasite (Bryant and Hallem, 2018a; Bryant and Hallem, 2018b; Bryant et al., 2018). Given the involvement of the TAX-2/TAX-4 channel in other sensory signaling pathways in free-living nematodes (Komatsu et al., 1996; Komatsu et al., 1999), it is likely that *S. stercoralis tax-4* knockouts will exhibit decrements in chemotactic as well as thermotactic behaviors when these mutants are subjected to the appropriate assays (Bryant and Hallem, 2018b; Bryant et al., 2018).

FUTURE DIRECTIONS FOR CRISPR/CAS9 IN PARASITIC NEMATODES

Deployment of CRISPR/Cas9 in Filariae

Robust systems for both integrative (Liu et al., 2018) and non-integrative (Xu et al., 2011) transgenesis are now available for *Brugia malayi*. With these systems in place, targeted mutagenesis and gene editing in this important model of human filariasis are imminently possible. The toolkit for integrative transgenesis assembled by Liu et al. (Liu et al., 2018) comprises a set of multifunctional vectors encoding multiple elements of the *piggyBac* transposon system that could be readily adapted to encode CRISPR/Cas9 elements such as the Cas9 endonuclease, gene-specific gRNAs, and templates for homology-directed repair in a compact and efficient gene delivery system. The time is right for this advancement.

Single-Copy Transgene Integrations

Among the very significant findings that arose from proof of principle for CRISPR/Cas9 in *Strongyloides* (Gang et al., 2017) was that full gene coding sequences, in this case an *Ss-act-2* transcriptional reporter encoding mRFPmars, can be integrated into the parasite genome at CRISPR/Cas9 cut sites by homology-directed repair. This approach provides a distinct advantage over current transposon-based methods for transgene integration, which integrate multiple transgene copies at random sites in the genome. Integrations by *piggyBac* typically favor coding sequences, creating the possibility of confounding insertional mutagenesis (Shao et al., 2012; Lok, 2013).

Regulatable CRISPR/Cas9

The capability to target transgene integrations to precise locations within the *Strongyloides* genome by CRISPR/Cas9 creates the potential to establish lines of these parasites that stably express the Cas9 endonuclease from single transgene copies. Parasites from such lines could subsequently be transformed to express gene-specific gRNAs to disrupt target genes and assess resulting phenotypes. Constitutive expression of Cas9 in *C. elegans* exerts some deleterious effects on fitness of the transgenic worms (Waaijers et al., 2013). Likewise, disrupting genes essential for invasion or establishment of the worms in mammalian hosts would prevent propagation of mutants and assessment of phenotypes affecting host infectivity, or crucial host–parasite interactions. A solution to these problems could be conditional or inducible Cas9 expression from integrated transgene sequences in stable parasite lines. We envision two possible approaches to regulating Cas9 expression in this context. The first would be to express Cas9 under promoters that are activated in the presence of a small molecule regulator such as tetracycline (Tet). Tet-responsive genes have been identified in *Brugia malayi* (Liu et al., 2011; Wurmthaler et al., 2019), and a tet-regulated promoter has been deployed for inducible transgene expression in *C. elegans* (Liu et al., 2011; Wurmthaler et al., 2019). These findings argue for identifying Tet-responsive genes in *Strongyloides* that could be similarly incorporated into **regulatable transgene expression**

systems generally and particularly for the regulated expression of Cas9 in stable parasite lines.

Another approach to regulatable expression of Cas9 and other transgene-encoded factors would be to fuse these to degradation domains that are stabilized in the presence of small molecules. Under these conditions, the transgene product would be stabilized in the presence of the small molecule but targeted to the proteasome for degradation in its absence. Such a system, incorporating the degradation domain of dihydrofolate reductase, which is stabilized in the presence of the folate-targeting drug trimethoprim, has been used to regulate transgene expression in *Plasmodium falciparum* (Muralidharan et al., 2011), and it has shown promise in preliminary experiments with transgenic Strongyloides (Lok et al., 2017).

TRANSGENES WITH DOMINANT MUTATIONS ARE TOOLS FOR FUNCTIONAL GENOMIC STUDY IN PARASITIC NEMATODES.

Functions of Genes Encoding a Variety of Catalytic or Regulatory Proteins may be Interrogated With Dominant Interfering Transgenes

Prior to the advent of targeted mutagenesis, transgenesis in *Strongyloides* spp. enabled a functional genomic approach based on expression of transgenes encoding mutant proteins designed to exert a dominant interfering effect on the target of interest. This approach may be used to suppress the functions of catalytic, signaling, and regulatory proteins that comprise separate functional domains and domains that bind to substrates, ligands, or DNA (**Figure 2**). The strategy holds that when overexpressed from multi-copy arrays relative to the single-copy endogenous target in the subject organism, transgene-encoded mutant proteins that have disrupted functional domains and intact binding domains will out-compete their endogenous wild-type counterparts for binding partners but fail to undertake their putative functions, thus resulting in dominant loss of function in the target gene of interest. Basic mechanisms for this approach to simulate loss of function in target genes are illustrated for cytoplasmic signaling kinases and metabolic enzymes (**Figure 2A**), membrane receptor kinases (**Figure 2B**), and transcription factors (**Figure 2C**).

A Dominant Transgene Approach has been Used to Investigate the Function of *Ss*-DAF-16 in *S. Stercoralis*

The gene encoding this transcription factor is the ortholog of *daf-16* in *C. elegans*, which encodes a forkhead transcription factor that is regulated by insulin-like signaling and controls lifespan and the switch between dauer and continuous larval development (Lin et al., 1997; Ogg et al., 1997; Lee et al., 2001; Massey et al., 2003; Massey et al., 2006). A loss-of-function approach to interrogating the function of *Ss-daf-16* employed a dominant
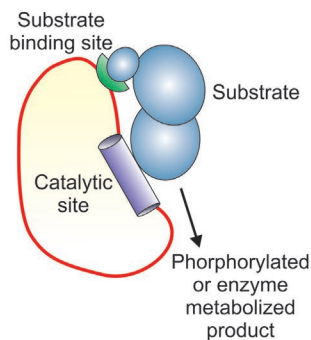
interfering construct. In this construct, AKT phosphorylation sites necessary for cytoplasmic localization of the protein were disrupted and the C-terminal domain of the molecule was truncated sufficiently to delete predicted transactivator binding sites (**Figure 2C**). Sequence encoding the DNA-binding domain of the mutant *Ss*-DAF-16 was left intact. The putative **dominant interfering transgene** also fused the coding sequence of the mutant *Ss-daf-16* with that of *gfp* to confirm constitutive nuclear localization of the encoded protein (Castelletto et al., 2009). This construct, therefore, was designed to express a mutant form of *Ss*-DAF-16 that is constitutively localized to the nucleus, but lacks the transactivating functions of the wild-type transcription factor. We also assume that the mutant transgene, like other plasmid-encoded transgenes introduced to nematodes by gonadal microinjection, will be incorporated into a multi-copy episomal array (Mello et al., 1991) and therefore overexpressed relative to the endogenous single-copy target gene (**Figure 2C**). Indeed, a proportion of F1 progeny of free-living female *S. stercoralis* microinjected with the mutant *Ss-daf-16* construct exhibited pronounced nuclear localization of GFP and a range of phenotypes including severe defects in intestinal morphology, loss of secretory granules from intestinal cells, retention of rhabditiform pharyngeal morphology in third-stage larvae (L3), and initiation of an aberrant L3–L4 molt (Castelletto et al., 2009). These last two phenotypes are particularly noteworthy given that 100% of wild-type progeny of free-living *S. stercoralis* females arrest their development as infective third-stage larvae, which have filariform, as opposed to rhabditiform morphology (Schad, 1989; Viney and Lok, 2015). In contrast to parasites expressing the putative dominant interfering construct, F1 progeny transformed with a control construct fusing the wild-type *Ss-daf-16* coding sequence to *gfp* exhibited normal intestinal architecture, with intestinal cells replete with secretory granules, filariform pharyngeal morphology in L3, and no evidence of L3–L4 molting (Castelletto et al., 2009).

A Dominant Interfering Transgene Construct was Used to Assess the Function of *Ss-riok-1* in *S. Stercoralis*

Ss-riok-1 encodes a RIO protein kinase homologous to RIOK-1 in *C. elegans* (Yuan et al., 2014b). RIO kinases are conserved in eukaryotic organisms and are essential for ribosomal biogenesis and cell cycle progression (Angermayr et al., 2002; Widmann et al., 2012; Yuan et al., 2014b). The putative dominant interfering construct encoded *Ss-riok-1* with a D282A mutation in its catalytic site that served to disrupt kinase activity. Sequence encoding the substrate binding site of *Ss-riok-1* was left intact in the putative dominant interfering construct (**Figure 2A**). As with the *Ss-daf-16* constructs described above, both the wild-type and mutant constructs encoding *Ss*-RIOK-1 fused its coding sequence to that of *gfp* to confirm an authentic pattern of expression under the *Ss-riok-1* promoter. As expected, both mutant and wild-type constructs were expressed in head and tail neurons and hypodermal cells of transgenic *S. stercoralis* larvae (Yuan et al., 2014b; Yuan et al., 2017). It was also expected that, by virtue of its situation in a multi-copy episomal array,
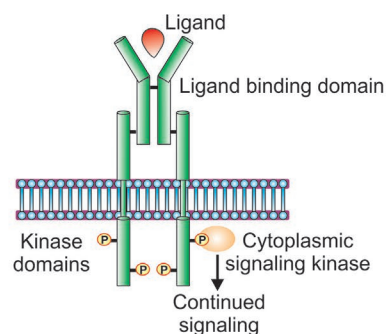
FIGURE 2 | Strategies for creating dominant interfering transgene constructs to study functions of catalytic, signaling, and gene regulatory proteins. **(A)**. Strategy for dominant interfering transgenes targeting functions of metabolic enzymes and cytoplasmic signaling kinases. Overexpression of mutant construct with intact substrate binding domain and catalytic site disrupted by mutation serves to outcompete endogenous gene product for substrate while yielding no phosphorylated or enzyme metabolized product. **(B)**. Design of dominant interfering transgenes targeting functions of membrane receptor kinases. Ablation of the kinase domain by mutation, an intact ligand binding domain and overexpression combine to out-compete the endogenous receptor kinase for ligand while failing to phosphorylate downstream cytoplasmic signaling elements. **(C)**. Design of dominant interfering transgenes targeting the action of transcription factors. Overexpression of a mutant transgene encoding an intact DNA binding domain, mutations in phosphorylation sites to effect constitutive nuclear localization, and disruption of the transactivating domain serve to outcompete the endogenous transcription factor for genomic response elements while failing to execute gene regulatory function.
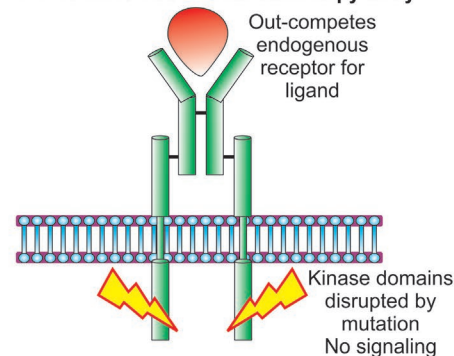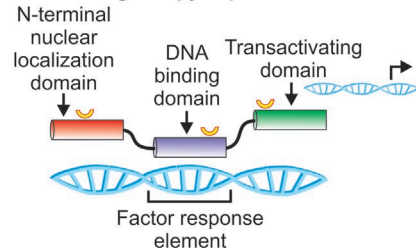
the mutant protein expressed from the dominant transgene construct would outcompete endogenous single-copy Ss-RIOK-1 for substrate binding but fail to exert its kinase function (**Figure 2A**). Larvae transformed with the construct encoding the kinase dead *Ss*-RIOK-1 exhibited profound decrements in motility and a significant blockade in development to the infective third-stage. By contrast, *S. stercoralis* larvae expressing a control construct encoding a fusion of *Ss*-RIOK-1 and GFP exhibited typical progressive motility and developed normally to the iL3 (Yuan et al., 2017). The study of *Ss-riok-1* function included an additional and crucial control experiment in which co-expression of wild-type *Ss*-RIOK-1 with the kinase-dead mutant served to rescue the motility and developmental phenotypes to a degree roughly proportional to concentration of wild-type vector plasmid microinjected into parental free-living female worms (Yuan et al., 2017). This important control experiment served to confirm the specificity of the dominant interfering transgene's effects on *Ss*-RIOK-1 function.

FUTURE DIRECTIONS IN THE USE OF DOMINANT TRANSGENE CONSTRUCTS TO STUDY GENE FUNCTION IN PARASITIC NEMATODES

Controlling Experiments Involving Dominant Interfering Constructs

Both studies employing dominant interfering constructs to evaluate specific gene function in *S. stercoralis* (Castelletto et al., 2009; Yuan et al., 2017) controlled for confounding effects of transgene overexpression by comparing phenotypes in cohorts of parasites overexpressing fusions of wild-type target proteins fused to GFP to those exhibited by parasites expressing mutant ones. This is crucial given the possibility that overexpression alone can, in some instances, produce dominant effects in transgenic organisms (Chandler and Werr, 2003). Another crucial control introduced in the study of *Ss*-RIOK-1 function was rescue of dominant interfering phenotypes by co-overexpression of the wild-type protein. In this case, the dose dependency of rescue was established by assessing frequencies of mutant phenotypes in F1 larvae derived from parental females microinjected with the mutant construct along with increasing concentrations of vector plasmid encoding the wild-type protein fused to GFP. There was a decreasing trend in the frequency of phenotypes with increasing concentration of wild-type vector plasmid, confirming the specificity of the mutant constructs dominant effect (Castelletto et al., 2009; Yuan et al., 2017). Inclusion of both of these essential controls will be crucial to the correct interpretation of data in future experiments with dominant interfering transgene constructs.

Deployment of Dominant Transgenes Incorporating Heterologous Activator or Suppressor Domains

To date, dominant interfering effects have been achieved with transgenes encoding the target gene with synthetic loss-of-function mutations in functional domains and wild-type sequence in DNA, protein, or substrate binding sites to facilitate competition for binding partners with the product of the endogenous target gene (**Figure 2**). Another approach to creating dominant interfering constructs targeting transcription factors in parasitic nematodes is to design the constructs to encode a chimeric protein comprising a heterologous transcriptional repressor domain fused to the DNA binding domain of the target factor. This serves to direct a strong repressor to promoters of genes regulated by the target transcription factor. An example of this approach is the use of the *Drosophila* ENGRAILED homeodomain to impart repressor function to the transcription factors, thereby creating loss-of-function phenotypes in both invertebrate and mammalian models (Badiani et al., 1994; John et al., 1995; Tolkunova et al., 1998; Chandler and Werr, 2003). Conversely, transgene constructs comprising viral activator domains such as VP16 from herpes simplex virus can activate target genes when fused to GAL-4 (Sevin-Pujol et al., 2017). The specific gene targeting capabilities of nucleoproteins comprising an inactive form of the Cas9 endonuclease, dCas9, and gRNAs can also be harnessed to tether repressor domains such as KRAB or activators such as VP64 to precise genome loci to achieve transcriptional repression (CRISPRi) or activation (CRISPRa) of target genes (Dominguez et al., 2016).

CONCLUSIONS

Transgenesis in *Strongyloides* and related parasitic nematodes in phylogenetic Clade IV and in the filaria *Brugia malayi* within Clade III opens possibilities for unprecedented studies of gene function in these pathogens that degrade the health of hundreds of millions of people. These approaches have enabled experiments employing dominant interfering transgenes to reveal functions of genes essential to ribosomal biogenesis and to morphogenesis and developmental arrest of infective larvae in *Strongyloides stercoralis*. Moreover, transgenesis has recently supported proof of principle for targeted mutagenesis *via* CRISPR/Cas9 in *Strongyloides* spp. Recent successes with integrative transgenesis in *Brugia malayi* create the possibility that similar studies of gene function will be possible in the filariae. Recent work with *Brugia* (Liu et al., 2018) and *Strongyloides* (Adams et al., 2019) illustrate the potential of lipofectant-mediated gene transfer, alone or in combination with microinjection, to facilitate applications of transgenesis and CRISPR/Cas9 gene disruption and editing to functional genomic study of a wide range of obligately parasitic nematodes with vector borne and free-living pre-parasitic larvae.

KEY CONCEPTS

Gonadal microinjection refers to a method of gene transfer whereby solutions of plasmid or linear transgene constructs are injected into the germlines of parental male or female worms. Nematode gonads frequently have syncytial regions that comprise germ cell nuclei within a common cytoplasm. Infusion of transgene construct solutions into these gonadal

syncytia results in significant numbers of transgenic oocytes or sperm.

Chemically mediated gene transfer, in the present context, involves gene transfer across the cuticles or exposed cellular surfaces of nematodes facilitated by incorporating transgene DNA into calcium phosphate co-precipitates or lipofectant micelles. Presenting these DNA preparations during intermolt periods of developing larvae, presumably when nascent cuticles are competent to take them up, may be crucial to this process.

Regulatable transgene expression is a strategy that allows transgene expression to be activated by the experimenter. This may be accomplished by using promoters that respond by activation to small molecules such as tetracycline or by coupling transgene encoded proteins of interest to degradation domains that are stabilized in the presence of some small molecule and degraded in the proteasome in its absence.

CRISPR/Cas9. Based on a prokaryotic defense response to foreign DNA, CRISPR/Cas9 harnesses sequence-specific targeting of the Cas9 endonuclease to precise gene loci by short "guide RNAs" to induce double-stranded DNA breaks. These may be repaired by non-homologous end joining, leaving random insertions or deletions, or by homology-directed repair, which can serve to incorporate a synthetic DNA repair template.

Dominant interfering transgenes encode mutations that ablate functional domains of a recombinant protein of interest, but leave crucial protein- or DNA-binding domains intact. When overexpressed, products of these dominant interfering transgenes outcompete the endogenous wild-type gene product for binding partners but fail to execute their designated functions, creating a loss-of-function phenotype in the subject.

AUTHOR CONTRIBUTIONS

JL formulated the concepts in the review, searched and analyzed the articles referenced, wrote all drafts of the text and designed and drafted the figure, which was then rendered by a graphic artist in the author's department and who is acknowledged in the paper.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Adams, S., Pathak, P., Shao, H., Lok, J. B., and Pires-daSilva, A. (2019). Liposome-based transfection enhances RNAi and CRISPR-mediated mutagenesis in non-model nematode systems. *Sci. Rep.* 9, 483. doi: 10.1038/s41598-018-37036-1

Angermayr, M., Roidl, A., and Bandlow, W. (2002). Yeast Rio1p is the founding member of a novel subfamily of protein serine kinases involved in the control of cell cycle progression. *Mol. Microbiol.* 44, 309–324. doi: 10.1046/j.1365-2958.2002.02881.x

Badiani, P., Corbella, P., Kioussis, D., Marvel, J., and Weston, K. (1994). Dominant interfering alleles define a role for c-Myb in T-cell development. *Genes Dev.* 8, 770–782. doi: 10.1101/gad.8.7.770

Bailey, M., Chauhan, C., Liu, C., and Unnasch, T. R. (2011). The role of polymorphisms in the spliced leader addition domain in determining promoter activity in *Brugia malayi*. *Mol Biochem. Parasitol.* 176, 37–41. doi: 10.1016/j.molbiopara.2010.11.012

Bargmann, C. I., and Mori, I. (1997). "Chemotaxis and Thermotaxis," in *C. elegans II*. Eds. Riddle, D. L., Blumenthal, T., Meyer, B. J. Priess, J. R. (Plainview, NY: Cold Spring Harbor Laboratory Press), 717–737.

Bouchery, T., Volpe, B., Shah, K., Lebon, L., Filbey, K., LeGros, G., et al. (2017). The study of host immune responses elicited by the model murine hookworms *Nippostrongylus brasiliensis* and *Heligmosomoides polygyrus*. *Curr. Protoc. Mouse Biol.* 7, 236–286. doi: 10.1002/cpmo.34

Bryant, A. S., and Hallem, E. A. (2018a). Temperature-dependent behaviors of parasitic helminths. *Neurosci. Lett.* 687, 290–303. doi: 10.1016/j.neulet.2018.10.023

Bryant, A. S., and Hallem, E. A. (2018b). Terror in the dirt: sensory determinants of host seeking in soil-transmitted mammalian-parasitic nematodes. *Int. J. Parasitol. Drugs Drug. Resist.* 8, 496–510. doi: 10.1016/j.ijpddr.2018.10.008

Bryant, A. S., Ruiz, F., Gang, S. S., Castelletto, M. L., Lopez, J. B., and Hallem, E. A. (2018). A critical role for thermosensation in host seeking by skin-penetrating nematodes. *Curr. Biol.* 28 (14), 2338–2347.e6. doi: 10.1016/j.cub.2018.05.063

Castelletto, M. L., Gang, S. S., Okubo, R. P., Tselikova, A. A., Nolan, T. J., Platzer, E. G., et al. (2014). Diverse host-seeking behaviors of skin-penetrating nematodes. *PLoS Pathog.* 10, e1004305. doi: 10.1371/journal.ppat.1004305

Castelletto, M. L., Massey, H. C., Jr., and Lok, J. B. (2009). Morphogenesis of *Strongyloides stercoralis* infective larvae requires the DAF-16 ortholog FKTF-1. *PLoS Pathog.* 5, e1000370. doi: 10.1371/journal.ppat.1000370

Chandler, J. W., and Werr, W. (2003). When negative is positive in functional genomics. *Trends Plant Sci.* 8, 279–285. doi: 10.1016/S1360-1385(03)00101-8

Davis, R. E., Parra, A., LoVerde, P. T., Ribeiro, E., Glorioso, G., and Hodgson, S. (1999). Transient expression of DNA and RNA in parasitic helminths by using particle bombardment. *Proc. Natl. Acad. Sci. U.S.A.* 96, 8687–8692. doi: 10.1073/pnas.96.15.8687

de Oliveira, A., Katholi, C. R., and Unnasch, T. R. (2008). Characterization of the promoter of the *Brugia malayi* 12kDa small subunit ribosomal protein (RPS12) gene. *Int. J. Parasitol.* 38, 1111–1119. doi: 10.1016/j.ijpara.2008.02.002

Dominguez, A. A., Lim, W. A., and Qi, L. S. (2016). Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. *Nat. Rev. Mol. Cell Biol.* 17, 5–15. doi: 10.1038/nrm.2015.2

Fire, A. (1986). Integrative transformation of *Caenorhabditis elegans*. *Embo J.* 5, 2673–2680. doi: 10.1002/j.1460-2075.1986.tb04550.x

Fire, A., Harrison, S. W., and Dixon, D. (1990a). A modular set of lacZ fusion vectors for studying gene expression in *Caenorhabditis elegans*. *Gene* 93, 189–198. doi: 10.1016/0378-1119(90)90224-F

Fire, A., Kondo, K., and Waterston, R. (1990b). Vectors for low copy transformation of C. elegans. *Nucleic. Acids Res.* 18, 4269–4270. doi: 10.1093/nar/18.14.4269

Gang, S. S., Castelletto, M. L., Bryant, A. S., Yang, E., Mancuso, N., Lopez, J. B., et al. (2017). Targeted mutagenesis in a human-parasitic nematode. *PLoS Pathog.* 13, e1006675. doi: 10.1371/journal.ppat.1006675

Grant, W. N., Skinner, S. J., Newton-Howes, J., Grant, K., Shuttleworth, G., Heath, D. D., et al. (2006a). Heritable transgenesis of *Parastrongyloides trichosuri*: a nematode parasite of mammals. *Int. J. Parasitol.* 36, 475–483. doi: 10.1016/j.ijpara.2005.12.002

Grant, W. N., Stasiuk, S., Newton-Howes, J., Ralston, M., Bisset, S. A., Heath, D. D., et al. (2006b). *Parastrongyloides trichosuri*, a nematode parasite of mammals that is uniquely suited to genetic analysis. *Int. J. Parasitol.* 36, 453–466. doi: 10.1016/j.ijpara.2005.11.009

Higazi, T. B., and Unnasch, T. R. (2004). Intron encoded sequences necessary for trans splicing in transiently transfected *Brugia malayi. Mol. Biochem. Parasitol.* 137, 181–184. doi: 10.1016/j.molbiopara.2004.04.014

Higazi, T. B., and Unnasch, T. R. (2013). Biolistic transformation of *Brugia malayi. Methods Mol. Biol.* 940, 103–115. doi: 10.1007/978-1-62703-110-3_9

Higazi, T. B., Merriweather, A., Shu, L., Davis, R., and Unnasch, T. R. (2002). *Brugia malayi*: transient transfection by microinjection and particle bombardment. *Exp. Parasitol.* 100, 95–102. doi: 10.1016/S0014-4894(02)00004-8

Higazi, T. B., Deoliveira, A., Katholi, C. R., Shu, L., Barchue, J., Lisanby, M., et al. (2005). Identification of elements essential for transcription in *Brugia malayi* promoters. *J. Mol. Biol.* 353, 1–13. doi: 10.1016/j.jmb.2005.08.014

Jackstadt, P., Wilm, T. P., Zahner, H., and Hobom, G. (1999). Transformation of nematodes *via* ballistic DNA transfer. *Mol. Biochem. Parasitol.* 103, 261–266. doi: 10.1016/S0166-6851(99)00089-4

John, A., Smith, S. T., and Jaynes, J. B. (1995). Inserting the Ftz homeodomain into engrailed creates a dominant transcriptional repressor that specifically turns off Ftz target genes *in vivo. Development* 121, 1801–1813.

Junio, A. B., Li, X., Massey, H. C., Jr., Nolan, T. J., Todd Lamitina, S., Sundaram, M. V., et al. (2008). *Strongyloides stercoralis*: cell- and tissue-specific transgene expression and co-transformation with vector constructs incorporating a common multifunctional 3′ UTR. *Exp. Parasitol.* 118, 253–265. doi: 10.1016/j.exppara.2007.08.018

Komatsu, H., Jin, Y. H., L'Etoile, N., Mori, I., Bargmann, C. I., Akaike, N., et al. (1999). Functional reconstitution of a heteromeric cyclic nucleotide-gated channel of *Caenorhabditis elegans* in cultured cells. *Brain. Res.* 821, 160–168. doi: 10.1016/S0006-8993(99)01111-7

Komatsu, H., Mori, I., Rhee, J. S., Akaike, N., and Ohshima, Y. (1996). Mutations in a cyclic nucleotide-gated channel lead to abnormal thermosensation and chemosensation in *C. elegans. Neuron* 17, 707–718. doi: 10.1016/S0896-6273(00)80202-0

Lee, R. Y., Hench, J., and Ruvkun, G. (2001). Regulation of C. elegans DAF-16 and its human ortholog FKHRL1 by the daf-2 insulin-like signaling pathway. *Curr. Biol.* 11, 1950–1957. doi: 10.1016/S0960-9822(01)00595-4

Lei, W. Q., Lok, J. B., Yuan, W., Zhang, Y. Z., Stoltzfus, J. D., Gasser, R. B., et al. (2017). Structural and developmental expression of *Ss-riok-2*, an RIO protein kinase encoding gene of *Strongyloides stercoralis. Sci. Rep.* 7, 8693. doi: 10.1038/s41598-017-07991-2

Li, X., Massey, H. C., Jr., Nolan, T. J., Schad, G. A., Kraus, K., Sundaram, M., et al. (2006). Successful transgenesis of the parasitic nematode *Strongyloides stercoralis* requires endogenous non-coding control elements. *Int. J. Parasitol.* 36, 671–679. doi: 10.1016/j.ijpara.2005.12.007

Li, X., Shao, H., Junio, A., Nolan, T. J., Massey, H. C., Jr., Pearce, E. J., et al. (2011). Transgenesis in the parasitic nematode *Strongyloides ratti. Mol. Biochem. Parasitol.* 179, 114–119. doi: 10.1016/j.molbiopara.2011.06.002

Lin, K., Dorman, J. B., Rodan, A., and Kenyon, C. (1997). *daf-16*: an HNF-3/forkhead family member that can function to double the life-span of *Caenorhabditis elegans.. Science* 278, 1319–1322. doi: 10.1126/science.278.5341.1319

Liu, C., Chauhan, C., and Unnasch, T. R. (2010a). The role of local secondary structure in the function of the trans-splicing motif of *Brugia malayi. Mol. Biochem. Parasitol.* 169, 115–119. doi: 10.1016/j.molbiopara.2009.10.003

Liu, C., Enright, T., Tzertzinis, G., and Unnasch, T. R. (2012). Identification of genes containing ecdysone response elements in the genome of *Brugia malayi. Mol. Biochem. Parasitol.* 186, 38–43. doi: 10.1016/j.molbiopara.2012.09.005

Liu, C., Kelen, P. V., Ghedin, E., Lustigman, S., and Unnasch, T. R. (2011). Analysis of transcriptional regulation of tetracycline responsive genes in *Brugia malayi. Mol. Biochem. Parasitol.* 180, 106–111. doi: 10.1016/j.molbiopara.2011.09.004

Liu, C., de Oliveira, A., Higazi, T. B., Ghedin, E., DePasse, J., and Unnasch, T. R. (2007). Sequences necessary for trans-splicing in transiently transfected *Brugia malayi. Mol. Biochem. Parasitol.* 156, 62–73. doi: 10.1016/j.molbiopara.2007.07.012

Liu, C., Oliveira, A., Chauhan, C., Ghedin, E., and Unnasch, T. R. (2010b). Functional analysis of putative operons in *Brugia malayi. Int. J. Parasitol.* 40, 63–71. doi: 10.1016/j.ijpara.2009.07.001

Liu, C., Mhashilkar, A. S., Chabanon, J., Xu, S., Lustigman, S., Adams, J. H., et al. (2018). Development of a toolkit for *piggyBac*-mediated integrative transfection of the human filarial parasite *Brugia malayi. PLoS Negl. Trop. Dis.* 12, e0006509. doi: 10.1371/journal.pntd.0006509

Liu, C., Voronin, D., Poole, C. B., Bachu, S., Rogers, M. B., Jin, J., et al. (2015). Functional analysis of microRNA activity in *Brugia malayi. Int. J. Parasitol.* 45, 579–583. doi: 10.1016/j.ijpara.2015.04.004

Lok, J. B. (2013). *piggyBac*: a vehicle for integrative DNA transformation of parasitic nematodes. *Mobile Genet. Elem.* 3, e24417. doi: 10.4161/mge.24417

Lok, J. B., and Massey, H. C., Jr. (2002). Transgene expression in *Strongyloides stercoralis* following gonadal microinjection of DNA constructs. *Mol Biochem. Parasitol.* 119, 279–284. doi: 10.1016/S0166-6851(01)00414-5

Lok, J. B., and Unnasch, T. R. (2013). "Transgenesis in animal parasitic nematodes: *Strongyloides* spp. and *Brugia* spp.," in *WormBook*, ed. The C. elegans Research Community (WormBook), 1–18. doi: 10.1895/wormbook.1.163.1

Lok, J. B., Shao, H., Massey, H. C., and Li, X. (2017). Transgenesis in *Strongyloides* and related parasitic nematodes: historical perspectives, current functional genomic applications and progress towards gene disruption and editing. *Parasitology* 144, 327–342. doi: 10.1017/S0031182016000391

Massey, H. C., Jr., Ranjit, N., Stoltzfus, J. D., and Lok, J. B. (2013). *Strongyloides stercoralis daf-2* encodes a divergent ortholog of *Caenorhabditis elegans* DAF-2. *Int. J. Parasitol.* 43, 515–520. doi: 10.1016/j.ijpara.2013.01.008

Massey, H. C., Bhopale, M. K., Li, X., Castelletto, M., and Lok, J. B. (2006). The fork head transcription factor FKTF-1b from *Strongyloides stercoralis* restores DAF-16 developmental function to mutant *Caenorhabditis elegans. Int. J. Parasitol.* 36, 347–352. doi: 10.1016/j.ijpara.2005.11.007

Massey, H. C., Jr., Nishi, M., Chaudhary, K., Pakpour, N., and Lok, J. B. (2003). Structure and developmental expression of *Strongyloides stercoralis fktf-1*, a proposed ortholog of *daf-16* in *Caenorhabditis elegans. Int. J. Parasitol.* 33, 1537–1544. doi: 10.1016/S0020-7519(03)00205-4

Mello, C., and Fire, A. (1995). DNA transformation. *Methods Cell Biol.* 48, 451–482. doi: 10.1016/S0091-679X(08)61399-0

Mello, C. C., Kramer, J. M., Stinchcomb, D., and Ambros, V. (1991). Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. *Embo J.* 10, 3959–3970. doi: 10.1002/j.1460-2075.1991.tb04966.x

Moerman, D. G., and Baillie, D. L. (1979). Genetic organization in *Caenorhabditis elegans*: fine-structure analysis of the unc-22 gene. *Genetics* 91, 95–103.

Moerman, D. G., Benian, G. M., Barstead, R. J., Schriefer, L. A., and Waterston, R. H. (1988). Identification and intracellular localization of the *unc-22* gene product of *Caenorhabditis elegans. Genes Dev.* 2, 93–105. doi: 10.1101/gad.2.1.93

Muralidharan, V., Oksman, A., Iwamoto, M., Wandless, T. J., and Goldberg, D. E. (2011). Asparagine repeat function in a *Plasmodium falciparum* protein assessed *via a* regulatable fluorescent affinity tag. *Proc. Natl. Acad. Sci. U.S.A.* 108, 4411–4416. doi: 10.1073/pnas.1018449108

Nolan, T. J., Megyeri, Z., Bhopale, V. M., and Schad, G. A. (1993). *Strongyloides stercoralis*: the first rodent model for uncomplicated and hyperinfective strongyloidiasis, the Mongolian gerbil (*Meriones unguiculatus*). *J. Infect. Dis.* 168, 1479–1484. doi: 10.1093/infdis/168.6.1479

Ogg, S., Paradis, S., Gottlieb, S., Patterson, G. I., Lee, L., Tissenbaum, H. A., et al. (1997). The fork head transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in *C. elegans. Nature* 389, 994–999. doi: 10.1038/40194

Schad, G. A. (1989). "Morphology and life history of *Strongyloides stercoralis*," in *Strongyloidiasis: a major roundworm infection of man*. Ed. D. I. Grove (Philadelphia: Taylor & Francis Inc.), 85–104.

Sevin-Pujol, A., Sicard, M., Rosenberg, C., Auriac, M. C., Lepage, A., Niebel, A., et al. (2017). Development of a GAL4-VP16/UAS trans-activation system for tissue specific expression in *Medicago truncatula. PLoS One* 12, e0188923. doi: 10.1371/journal.pone.0188923

Shao, H., Li, X., and Lok, J. B. (2017). Heritable genetic transformation of *Strongyloides stercoralis* by microinjection of plasmid DNA constructs into the male germline. *Int. J. Parasitol.* 9, 511–515. doi: 10.1016/j.ijpara.2017.04.003

Shao, H., Li, X., Nolan, T. J., Massey, H. C., Jr., Pearce, E. J., and Lok, J. B. (2012). Transposon-mediated chromosomal integration of transgenes in the parasitic nematode *Strongyloides ratti* and establishment of stable transgenic lines. *PLoS Pathog.* 8, e1002871. doi: 10.1371/journal.ppat.1002871

Shu, L., Katholi, C. R., Higazi, T., and Unnasch, T. R. (2003). Analysis of the *Brugia malayi* HSP70 promoter using a homologous transient transfection system. *Mol Biochem. Parasitol.* 128, 67–75. doi: 10.1016/S0166-6851(03)00052-5

Stoltzfus, J. D., Massey, H. C., Jr., Nolan, T. J., Griffith, S. D., and Lok, J. B. (2012). *Strongyloides stercoralis age-1*: a potential regulator of infective larval development in a parasitic nematode. *PLoS One* 7, e38587. doi: 10.1371/journal.pone.0038587

Tolkunova, E. N., Fujioka, M., Kobayashi, M., Deka, D., and Jaynes, J. B. (1998). Two distinct types of repression domain in engrailed: one interacts with the groucho corepressor and is preferentially active on integrated target genes. *Mol. Cell Biol.* 18, 2804–2814. doi: 10.1128/MCB.18.5.2804

Tzertzinis, G., Egana, A. L., Palli, S. R., Robinson-Rechavi, M., Gissendanner, C. R., Liu, C., et al. (2010). Molecular evidence for a functional ecdysone signaling system in *Brugia malayi*. *PLoS Negl. Trop. Dis.* 4, e625. doi: 10.1371/journal.pntd.0000625

Viney, M. E., and Lok, J. B. (2015). "The biology of Strongyloides spp.," in *WormBook*, ed. The C. elegans Research Community (WormBook), 1–17. doi: 10.1895/wormbook.1.141.2

Waaijers, S., Portegijs, V., Kerver, J., Lemmens, B. B., Tijsterman, M., van den Heuvel, S., et al. (2013). CRISPR/Cas9-targeted mutagenesis in *Caenorhabditis elegans*. *Genetics* 195, 1187–1191. doi: 10.1534/genetics.113.156299

Widmann, B., Wandrey, F., Badertscher, L., Wyler, E., Pfannstiel, J., Zemp, I., et al. (2012). The kinase activity of human Rio1 is required for final steps of cytoplasmic maturation of 40S subunits. *Mol. Biol. Cell* 23, 22–35. doi: 10.1091/mbc.e11-07-0639

Wurmthaler, L. A., Sack, M., Gense, K., Hartig, J. S., and Gamerdinger, M. (2019). A tetracycline-dependent ribozyme switch allows conditional induction of gene expression in *Caenorhabditis elegans*. *Nat. Commun.* 10, 491. doi: 10.1038/s41467-019-08412-w

Xu, S., Liu, C., Tzertzinis, G., Ghedin, E., Evans, C. C., Kaplan, R., et al. (2011). *In vivo* transfection of developmentally competent *Brugia malayi* infective larvae. *Int. J. Parasitol.* 41, 355–362. doi: 10.1016/j.ijpara.2010.10.005

Yuan, W., Liu, Y., Lok, J. B., Stoltzfus, J. D., Gasser, R. B., Lei, W., et al. (2014a). Exploring features and function of *Ss-riok-3*, an enigmatic kinase gene from *Strongyloides stercoralis*. *Parasit. Vectors* 7, 561. doi: 10.1186/PREACCEPT-1041797881140939

Yuan, W., Lok, J. B., Stoltzfus, J. D., Gasser, R. B., Fang, F., Lei, W. Q., et al. (2014b). Toward understanding the functional role of *Ss-RIOK-1*, a RIO protein kinase-encoding gene of *Strongyloides stercoralis*. *PLoS Negl. Trop. Dis.* 8, e3062. doi: 10.1371/journal.pntd.0003062

Yuan, W., Zhou, H., Lok, J. B., Lei, W., He, S., Gasser, R. B., et al. (2017). Functional genomic exploration reveals that *Ss-RIOK-1* is essential for the development and survival of *Strongyloides stercoralis* larvae. *Int. J. Parasitol.* 47, 933–940. doi: 10.1016/j.ijpara.2017.06.005

frontiers
in Genetics

Omics Driven Understanding of the Intestines of Parasitic Nematodes

Douglas P. Jasmer[1]*, Bruce A. Rosa[2], Rahul Tyagi[2] and Makedonka Mitreva[2,3]*

[1] Department of Veterinary Microbiology and Pathology, Washington State University, Pullman, WA, United States,
[2] McDonnell Genome Institute, Washington University, St. Louis, St. Louis, MI, United States [3] Department of Internal Medicine, Washington University School of Medicine, St. Louis, MI, United States

The biological and molecular complexity of nematodes has impeded research on development of new therapies for treatment and control. We have focused on the versatility of the nematode intestine as a target for new therapies. To that end, it is desirable to establish a broad and deep understanding of the molecular architecture underlying intestinal cell functions at the pan-Nematoda level. Multiomics data were generated to uncover the evolutionary principles underlying both conserved and adaptable features of the nematode intestine. Whole genomes were used to reveal the functional potential of the nematodes, tissue-specific transcriptomes provided a deep assessment of genes that are expressed in the adult nematode intestine, and comparison of selected core species was used to determine a first approximation of the pan-Nematoda intestinal transcriptome. Differentially expressed transcripts were also identified among intestinal regions, with the largest number expressed at significantly higher levels in the anterior region, identifying this region as the most functionally unique compared to middle and posterior regions. Profiling intestinal miRNAs targeting these genes identified the conserved intestinal miRNAs. Proteomics of intestinal cell compartments assigned proteins to several different intestinal cell compartments (intestinal tissue, the integral and peripheral intestinal membranes, and the intestinal lumen). Finally, advanced bioinformatic approaches were used to predict intestinal cell functional categories of seminal importance to parasite survival, which can now be experimentally tested and validated. The data provide the most comprehensive compilation of constitutively and differentially expressed genes, predicted gene regulators, and proteins of the nematode intestine. The information provides knowledge that is essential to understand molecular features of nematode intestinal cells and functions of fundamental importance to the intestine of many, if not all, parasitic nematodes.

Keywords: nematode, intestine, genome, transcriptome, proteome, miRNA, dsRNA

INTRODUCTION

Infections caused by parasitic nematodes result in substantial mortality and morbidity in human populations, especially in tropical regions of Africa, Asia, and the Americas (Pullan et al., 2014). Infections by parasitic nematodes also compromise the health and productivity of livestock species on a global basis. The impact on food production most significantly affects human health in underdeveloped regions of the world where small holders of livestock depend on food animals for basic nutrition and commerce (Perry and Grace, 2009). Losses in livestock production directly deplete resources needed

for basic nutrition, income, and trade. The negative impact of parasitic nematode infections in humans and animals is reduced by anthelmintic treatments; however, the propensity of these pathogens to acquire resistance to anthelmintics (Kaplan, 2004; Leathwick, 2013) threatens the existing methods to prevent and treat diseases they cause. Hence, there is a continuing and compelling need to identify new approaches, therapeutic targets, and applications to prevent and treat infections by nematode pathogens.

This review summarizes the progress made in the last 15 years on an approach that focuses on a single tissue of importance for survival of parasitic nematodes (the intestine) with a goal of developing an integrated research model that will have pan-Nematoda application for deriving essential and broadly conserved intestinal cell functions of these pathogens. One anticipated application of this knowledge is advancement toward the development of pharmaco- and immunotherapeutics for the treatment, prevention, and control of infections caused by nematode pathogens. Earlier studies have validated the value of the intestine of parasitic nematodes for this purpose (next section), which prompts the need to develop resources that can accelerate research on this tissue. The summarized progress relates to the establishment of a multiomics database approach to identify pan-Nematoda conserved intestinal cell functions, coupled with the development of methodologies that can be used to experimentally dissect those conserved functions with new applications to therapeutics as one end goal.
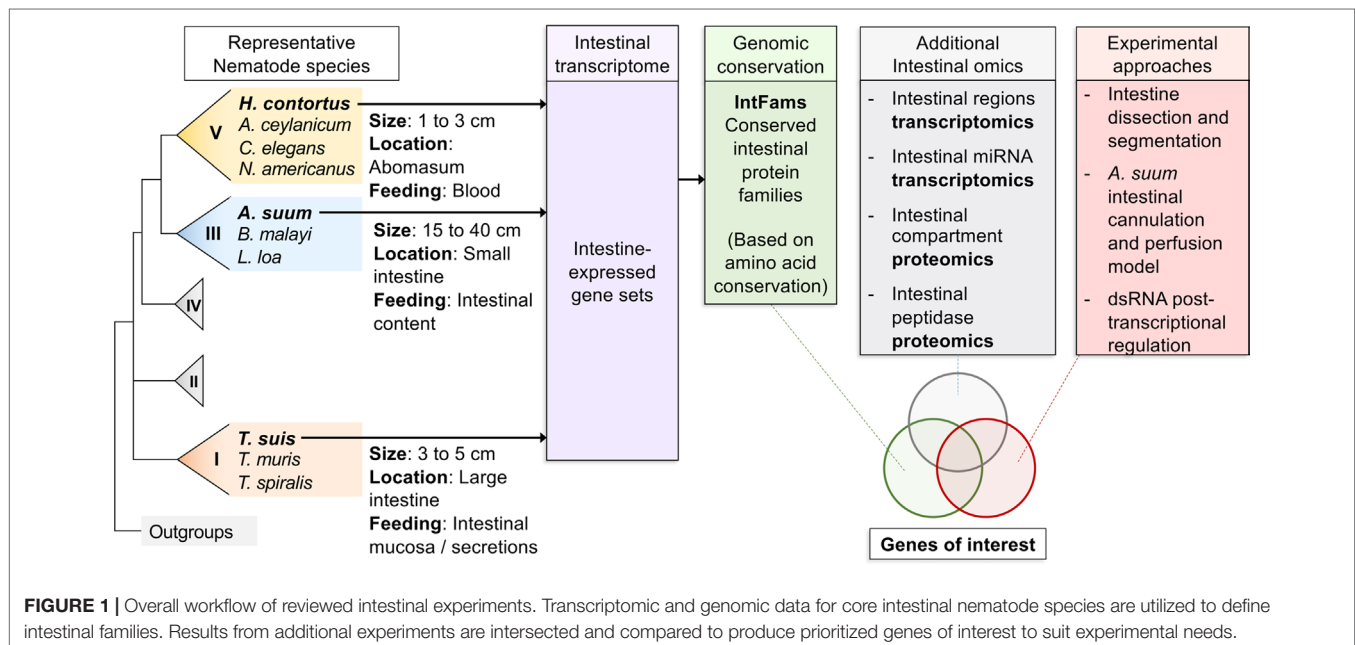
THE NEMATODE INTESTINE AS A TARGET FOR NEW ANTHELMINTIC THERAPIES
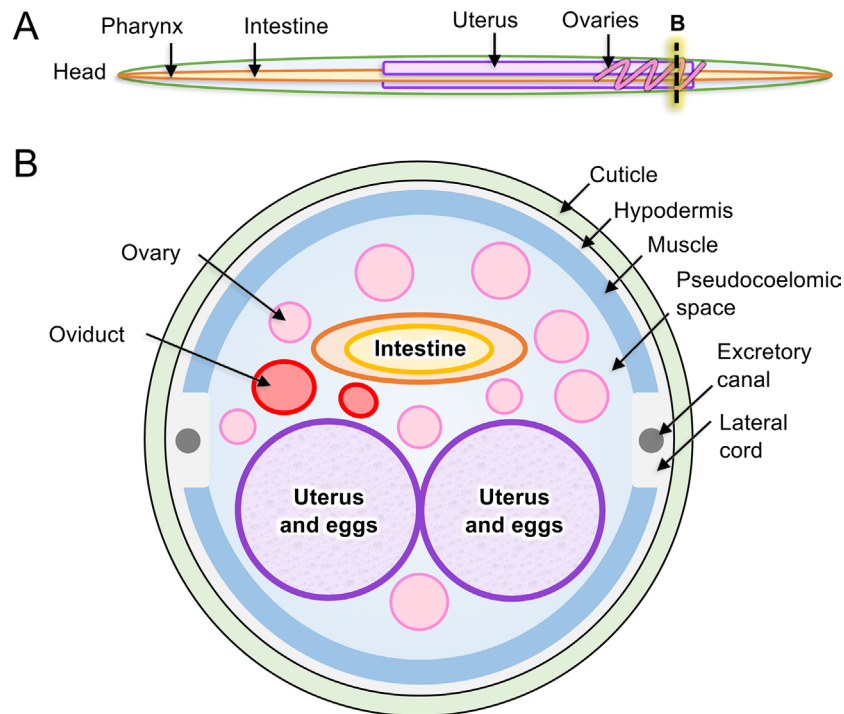
General Considerations

Multiple approaches can be considered when aiming to identify targets for pharmacotherapy, including targets that are expressed in all tissues and parasitic stages and conserved across diverse nematode pathogens (**Figure 1**). However, this requirement overlooks the idiosyncrasies of individual parasite tissues, especially since many, although not all, marketed anthelmintics appear to specifically target one system, the neuromuscular system. Similarly, other tissue/organ systems of nematodes, if disrupted, can theoretically offer the same value for pharmacological targets as the neuromuscular system. The nematode intestine is but one example where experimental data support this view, as will be described in this section. One challenge to more forcefully move forward on tissue-specific research is the relatively shallow knowledge on specific functions, and then molecules, that perform or regulate basic functions in tissues of parasitic nematodes. Furthermore, a focus on an individual tissue does not exclude the possibility that findings generated from one tissue will have application to many or all other tissues and/or stages of parasitic nematodes.

The tubular intestine of nematodes is composed of polarized epithelial-like cells arranged as a single cell layer (**Figure 2**). Some parasitic nematodes, such as the filarial nematodes, have a somewhat vestigial intestine, possibly due to relocating elements of nutrient acquisition to the cuticular surface (Howells and Chen, 1981). Nevertheless, more recent evidence has clarified roles that intestinal cells of filariae appear to perform (Morris et al., 2015; Ballesteros et al., 2018). The apical intestinal membrane (AIM) of the nematode intestine is continuous with the external environment *via* the nematode stoma and, as such, represents in parasitic species an internal interface with the host. Numerous essential functions are known, or expected, to exist on the AIM related to nutrient digestion and acquisition, general homeostasis of intestinal cells and the intestinal tract, and protection from toxins. Additionally, regeneration of damaged tissue *via* cell proliferation is not known to occur in nematodes. In the absence of effective regeneration, which

FIGURE 1 | Overall workflow of reviewed intestinal experiments. Transcriptomic and genomic data for core intestinal nematode species are utilized to define intestinal families. Results from additional experiments are intersected and compared to produce prioritized genes of interest to suit experimental needs.

FIGURE 2 | Anatomy of model intestinal nematode species *Ascaris suum*. **(A)** Major organs are identified in an adult *A. suum* female worm, including **(B)** in cross-section.

still requires additional investigation to validate, therapies that cause intestinal cell death and breach of the single cell layer separating the lumen from the pseudocoelom are likely to have lethal outcomes for the parasite. Several lines of research indicate that the intestine of parasitic nematodes has high potential for targeting in immunotherapies and pharmacotherapies that are relevant to treatment, prevention, and control of infections by parasitic nematodes as discussed here.

Immunotherapeutic Applications of Nematode Intestinal Antigens

From an immunotherapeutic perspective, digestive enzymes localized to the AIM provide one example of antigens for inducing protective immune responses, and antibodies in particular. The expectation is that when the relevant antibodies are ingested by parasitic nematodes, they can inhibit nutrient digestion leading to parasite expulsion. Progress on this approach has been most impressive in blood-feeding parasitic nematodes, such as *Haemonchus contortus* and hookworms, leading to vaccine preparations with potential value (Knox et al., 2003; Knox, 2011; Hotez et al., 2013; Bassetto et al., 2018). However, application of this approach to parasitic nematodes that feed at other host locations has not been sufficiently investigated. For instance, *Ascaris* spp. live in the lumen of the small intestine, and *Trichuris* spp. burrow into the lining of the

colon and cecum. Additionally, it remains to be determined if the immune mechanism(s) involved depends on inhibition of functions within the intestinal lumen of the parasite, which has been an underlying premise of this approach. Intestinal antigens from *H. contortus* also localize to the abomasal mucosa and induce local mucosal responses during infection (Jasmer et al., 1996; Jasmer et al., 2007), the importance of which to protective immunity remains undetermined. Despite lack of knowledge of mechanisms involved, intestinal antigens have proven value for inducing immunity against some parasitic nematodes and have likely application to others.

A more recent nuance of possible immunization with intestinal antigens involves exosomes from *Heligmosomoides polygyrus* that are isolated from excretory–secretory products. The exosomes are thought to derive at least in part from the intestine of *H. polygyrus* released in intestinal excretions from the worm (Buck et al., 2014). Immunization with the exosome preparation induced protective immunity against challenge infections in mice (Coakley et al., 2017). It will be of interest to learn if there is a role of intestinal antigens in inducing this immunity.

Pharmacotherapies That Cause Irreparable Damage to the Nematode Intestine

Effective pharmacotherapy that targets the nematode intestine has previously been demonstrated. The first and most thoroughly

investigated examples involve benzimidazole anthelmintics. Benzimidazole treatment of host animals infected with *Ascaris suum*, *H. contortus*, or other parasitic nematodes leads to disruption of intestinal cell architecture and frank disintegration of intestinal tissue; the range of observations depends on species and experimental designs (Borgers and De Nollin, 1975; Borgers et al., 1975; Atkinson et al., 1980; Comley, 1980; Zintz and Frank, 1982; Jasmer et al., 2000; Hanser et al., 2003; O'Neill et al., 2015). The anterior region of the *H. contortus* intestine displayed hypersensitivity regarding these effects, which reflected the occurrence of relatively rapid (within 12 h), irreparable intestinal cell damage, inclusive of nuclear DNA fragmentation, from which the parasite is unlikely to recover (Jasmer et al., 2000). Meanwhile, in *Brugia malayi*, flubendazole exposure does significantly less damage to the intestine, and after a brief exposure *in vitro*, the intestinal and hypodermal damage to adult worms resolved over time after transplantation into the host (O'Neill et al., 2016). The observations clearly show that idiosyncrasies of specific tissues and tissue regions are of importance in the context of inducing irreparable tissue damage, which has high value as an end point for effective anthelmintics. Regarding a possible mechanism, current evidence supports that benzimidazoles inhibit microtubule-mediated apical vesicle transport of hydrolytic enzymes (digestive enzymes) in intestinal cells, which then become dispersed in the cytoplasm where intracytoplasmic digestion of intestinal cells/tissue might occur (Shompole et al., 2002). Stating this possibility in another way, benzimidazoles appear to induce a pathologic process that leads to irreparable intestinal cell damage, which goes beyond simple disruption of nutrient acquisition as might be expected for a downstream effect on intestinal cells by neuromuscular toxins that inhibit pharyngeal pumping. It seems clear that elucidating the actual mechanism responsible will present opportunities to identify alternative methods (targets and drugs) to induce this irreparable damage. If the specific hypothesis is correct, then multiple cellular components of the apical secretory process represent potential targets for inducing this lethal effect in nematode intestinal cells. An additional possibility is that other pathologic processes contribute to, or are fully responsible for, the irreparable damage to intestinal cells. Hence, it will be important to develop methods that can distinguish among these possible explanations. The pan-Nematoda multiomics resource described below can greatly facilitate progress on sorting these alternative explanations and identification of potential new anthelmintic targets.

A second example involves use of suramin against filarial nematodes (Howells et al., 1983) in which the anthelmintic effect was attributed to activity against the intestine, producing ultrastructural changes. Notably, the timeframe for efficacy extended to 5 weeks posttreatment, indicating survival for a prolonged period despite apparent intestinal tissue damage.

A third example of pharmacotherapy targeting of the intestine involves the use of *Bacillus thuringiensis* crystal (CRY) protein toxins. The pore-forming CRY toxins bind to the AIM upon ingestion by insects and nematodes and initiate an irreparable pathologic process that leads to death of the insect or nematode, presumably resulting from generation of pores in the AIM (Hu and Aroian, 2012). This effect also appears (Huffman et al., 2004; Hu and Aroian, 2012) to involve only intestinal cells in the live

invertebrates. CRY proteins have also demonstrated interesting utility as anthelmintics for parasitic nematode infections when administered to host animals (Cappello et al., 2006). Once again, it is possible that other inducible pathologic processes contribute to the demise of intestinal cells initiated by CRY toxin. For instance, intestinal factors (p38 MAP kinase pathway) that confer protection against CRY toxins have been identified in *C. elegans*, and better understanding of the protective mechanisms could aid understanding of pathologic mechanisms involved.

A fourth example involves trans-cinnamaldehyde-induced destruction of intestinal tissue in *A. suum* larval stages (Williams et al., 2015). It is worth noting that the pathologic mechanism(s) induced is unclear, and the damage is not restricted to intestinal cells; nonetheless, prior studies have contributed to our knowledge on the susceptibility of intestinal cells in parasitic nematodes to diverse anthelmintic modalities.

Take-Home Lessons From Past Findings

Major points that can be taken from the foregoing examples include: 1) biological idiosyncrasies (cellular processes specific to the intestine, or accessibility by location on an internal interface with the host) make the intestine of parasitic nematodes a particularly compelling target for immunotherapies and pharmacotherapies; 2) components of the AIM are implicated in roles related to mechanisms of action for at least three of the five (AIM antigens, benzimidazole anthelmintics, and CRY toxins) described immunotherapeutic and pharmacotherapeutic examples, which makes this membrane surface especially interesting; 3) specific processes either involving inhibition of apical secretion or disruption of the AIM have been implicated in pathology induced by benzimidazole anthelmintics and CRY toxin anthelmintics, respectively; and 4) idiosyncrasies of the intestine translate into apparent irreparable tissue damage for at least two of the four (benzimidazoles and CRY toxins) pharmacotherapy examples discussed. The various immunological and pharmacological findings have different conceptual origins (e.g., direct inhibition of digestion and nutrient acquisition, inhibition of apical secretion, or direct disruption of the AIM, respectively), which might lead to an array of new targets to achieve similar outcomes once the mechanisms are better understood for each of them. Hence, there are compelling reasons to improve capabilities for research on intestinal cells in parasitic nematodes, including generation of deeper molecular understanding of nematode intestinal cell functions and developing experimental capabilities to test predictions, progress on both of which are described in the sections below.

PAN-NEMATODA AND MULTIOMICS APPROACH TO RESEARCH ON THE INTESTINE

General Rationale

Despite most nematodes having a recognizable intestinal tract, the morphology and organization of this tissue are diverse among groups, suggesting functional diversity. At a tissue level,

intestinal development ranges from superficial or somewhat vestigial (Eisenback, 1985), to the more common occurrence of a fully developed intestinal tract (Munn and Greenwood, 1984). Nevertheless, intestinal cells reportedly can be organized as syncytia (multinucleate, Munn and Greenwood, 1984), or not, with polyploid nuclei that vary among species from an estimated 4N for adult *A. suum* (Anisimov et al., 1975) to 32N for adult *C. elegans* intestinal cells (Hedgecock and White, 1985). Otherwise, phylogenetic variation of intestinal cells has generally been understudied. More recent progress is summarized in the following sections to clarify basic intestinal cell functions that are conserved at molecular and cellular levels among most nematode species. This progress was facilitated by the selection of core species that span the phylum and represent much of the phylogenetic distance across the Nematoda. The design supported development of a pan-Nematoda multiomics database that was interrogated to investigate broadly conserved biological functions of intestinal cells/tissue for applications toward new pharmacotherapies and immunotherapies.

Selection of Research Core Species That Sample the Broad Phylogenetic Diversity of the Nematoda

Core adult parasite species have been selected based on four main criteria: 1) each represented a major lineage that collectively represented much of the phylogenetic diversity of the Nematoda and concurrently incorporated the fewest species needed for this purpose; 2) each species could easily be dissected to provide sufficient intestinal tissue for high quality transcriptomic data generation and analyses; 3) extensive genome sequence was available, or near-term obtainable; and 4) each species presented potential for development, or was already established, as an experimental model for research on the intestine. The core species selected (**Table 1**) included: 1) *H. contortus*, a clade V haematophagous parasitic nematode of small ruminants and longtime model for research on its intestine; 2) *A. suum*, a clade III parasitic nematode of the small intestine of swine and humans, the large size of which provided unique advantages that facilitated this research; and 3) *Trichuris suis*, a clade I parasitic nematode of the cecum and large intestine of swine that supported our criteria and provided a needed connection to the most ancient clade of the Nematoda. Each of these core species also has importance

for human or veterinary medicine and provides examples of soil-transmitted helminths in their own context. Finally, each also occupies distinct anatomical locations in the host and obtains nutrients from distinct host compartments (blood, intestinal lumen content/epithelial cells, or nutrients accessible from attachment to the mucosa, respectively). Thus, these three core species represent the extensive phylogenetic and biologic diversity sought with this design.

The multiomics approach delineated the elements of genomes, transcriptomes, proteomes (predicted or/and determined by mass spectrometry), and regulomes (microRNA) that serve intestinal cell functions in a pan-Nematoda context. This effort identified intestinal genes, predicted RNAs, microRNAs, and proteins for which degree of conservation among species was documented to the extent possible. Resulting knowledge bases generated are summarized in the sections below and in **Table 2**.

ANTECEDENTS TO DEEP PAN-NEMATODA INTESTINAL TRANSCRIPTOMICS DATABASE DEVELOPMENT

Genome Resources

Next generation sequencing technologies revolutionized many biomedical research fields including parasitology. As technology was advancing, genomes and predicted proteome sequences for nematode species were accumulating and were subsequently improved over the past two decades. The efforts resulted in generating low coverage genome survey sequences and draft genomes before good quality assemblies and annotations were available. While great progress has been made (Coghlan et al., 2019), most of the available parasitic nematode genomes are draft assemblies with many gaps to be closed and accuracy to be improved (with only a few exceptions, such as *H. contortus* and *O. volvulus*) (Cotton et al., 2016; Laing et al., 2016). Postgenomic applications frequently require comparative genomics on a gene and single nucleotide level, and performing these analyses on draft genomes is inadequate. It is more than ever important to continue efforts on minimizing gene fragmentations, eliminating gene model errors, and resolving collapse of recently duplicated and diverged sequences.

TABLE 1 | Genome statistics for core helminth species used to study intestinal functions.

Statistic	*Haemonchus contortus* (Howe et al., 2017)*	*Trichuris suis* (Coghlan et al., 2019)	*Ascaris suum* (original) (Jex et al., 2011)	*Ascaris suum* (improved) (Wang et al., 2017)
Genome size (Mb)	283.4	63.8	272.8	298.0
Coding genes	19,430	9,831	18,542	16,778
Contigs	192	5,498	46,119	1,527
Scaffolds	7	306	29,831	415
BUSCO completeness	92.2%	80.8%	96.4%	96.5%
N50 length (Mb)	47.4	1.3	0.41	4.6
N50 number	3	16	179	21

*Improved version of published genome (Laing et al., 2013).

TABLE 2 | Available omics resources useful for the study of the helminth intestine.

Dataset type	Dataset description	Use for intestinal studies
Genomic	Genome assembly/annotation for *A. suum* (Jex et al., 2011), *H. contortus* (Laing et al., 2013), and *T. suis* (Wang et al., 2015)	Source for all gene and protein sequence information for core helminth intestine species
	Ortholog matches to other species (Rosa et al., 2014; Wang et al., 2015)	Gene conservation across helminth species/helminth-specific genes
	Wellcome Sanger Institute genome resource (www.sanger.ac.uk/resources/downloads/helminths)	Additional helminth genomes and BLAST database
	Wormbase Parasite (https://parasite.wormbase.org) (Howe et al., 2017)	Helminth genome repository/BLAST services, gene enrichment testing/variant effect predictor
Genomic functional annotations	KEGG (Kanehisa et al., 2016)	Biological pathways relevant to intestine
	Gene ontology (Ashburner et al., 2000)	GO terms related to intestine
	MEROPS (peptidases) (Rawlings and Morton, 2008)	Specific classification of digestive peptidases
	Interpro domains (Hunter et al., 2012; Jones et al., 2014)	Additional specific functional annotation
	5′ upstream UTR binding motifs (Rosa et al., 2014)	Identify transcription factors that may modulate expression of intestinal genes
	RNAi phenotype in *C. elegans* (Piano et al., 2000; Piano et al., 2002; Fernandez et al., 2005; Sönnichsen et al., 2005)	Identify intestine-related and lethal/sterile phenotypes
	Protein–protein interactions* (Simonis et al., 2009)	Target intestinal proteins that interact with many other intestinal proteins
Transcriptomic	RNA-Seq of 10 *A. suum* tissues (Rosa et al., 2014)	Identify intestine-overexpressed genes in adult male and female *A. suum*
	RNA-Seq of anterior, middle and posterior *A. suum* intestine (Gao et al., 2016)	Identify intestinal genes and miRNAs expressed more highly in various regions of the intestine
	Intestinal RNA-Seq expression from *H. contortus* and *T. suis* (Wang et al., 2015)	Identify intestine-expressed genes in phylogenetically distinct *A. suum* orthologs
Proteomic	Proteins detected in various A. suum intestinal compartments (Rosa et al., 2015)	Confident identification of proteins from intestinal tissue, apical and peripheral intestinal membranes, intestinal lumen, and pseudocoelomic fluid
	Peptidases in *A. suum* intestine (Jasmer et al., 2015)	Detailed list of proteomics-confirmed peptidases on the apical intestinal membrane, and in the intestinal lumen
Multiomics	Intestinal families (IntFams) (Wang et al., 2015)	Confident nematode-conserved and consistently intestine-expressed genes using multiple species
	Helminth.net online resource (Martin et al., 2015)	A collection of omics databases with tools to search genes and functions, perform BLAST searches, view KEGG pathways, browse variants, and perform multiomics comparisons to identify drug and vaccine targets

Inferred using the closest significant C. elegans ortholog.

Early Intestinal Transcriptome Research

Experimental observations listed in the previous section motivated early efforts to expand knowledge of intestinal proteins that are expressed by individual parasitic nematodes (Gao et al., 2016), those that are conserved among nematodes (Wang et al., 2015), and those that differentiate intestinal versus other tissues in nematodes (Rosa et al., 2014). Initial comparisons utilized expressed sequence tag (EST) (Yin et al., 2008) and gene microarray (Wang et al., 2013b) methods, and identified a modest number of predicted intestinal protein families (IntFams) that are orthologs conserved among *H. contortus*, *A. suum*, and *C. elegans*, as well as major differences. One notable difference involved a large family of intestinal cathepsin B-like cysteine proteases expressed by *H. contortus*, whereas a single intestinal family member was detected for *A. suum* and 12 family members have been described for *C. elegans* (Jasmer et al., 2004). This early foundation fostered anticipation that while many intestinal

functions may be conserved among nematodes, there are likely many adaptations that differentiate nematodes species, and possibly phylogenetic lines.

Due to its large size, *A. suum* (male, 15–30 cm; female, 20–35 cm) is a particularly good nematode model to study tissue transcriptomics, since the individual tissues can be cleanly and accurately dissected. In the first ever comprehensive tissue-specific RNAseq-based transcriptome studies for any parasitic nematode, comparative transcriptomics was performed on three nonreproductive tissues (head, pharynx, and intestine) in both male and female worms, as well as four reproductive tissues (testis, seminal vesicle, ovary, and uterus) (Rosa et al., 2014). This study identified thousands of genes associated with the different tissues (or combinations of tissues), including 1,387 genes overexpressed in the intestine samples (male or female). Examination of the 5′ untranslated region (UTR) sequences preceding the coding sequences for the intestine-associated genes identified

enrichment for the motif that is bound by the transcription factor ELT-2, the predominant transcription factor controlling intestinal development and function in *C. elegans* (McGhee et al., 2009). Intestine-enriched functions included nine "Molecular Function" child terms of "hydrolase activity" (GO:0016787; including "cysteine-type endopeptidase activity"). Gene coexpression networks linked some genes of the highly intestine overexpressed genes to other tissues, including the head-overexpressed FAR-1 ortholog, a proposed anthelmintic target with a crucial role in parasitism (Bradley et al., 2001; Basavaraju et al., 2003). Overall, this study provided a foundation for cataloguing and profiling intestinal expression in *A. suum*, and a gene expression and annotation database was built using its datasets (Rosa et al., 2014).

Pan-Nematoda Database of Intestine Expressed Gene Transcripts and Predicted Proteins

The next step after the *A. suum* comparative tissue expression data was to determine the level of conservation of intestinal cell functions among nematode lineages and species. Three omics resources were central to this investigation. The first was annotated genomes and predicted proteome information for each of the core species under investigation. The second was deep intestinal transcriptomes (generated by RNAseq) for each of the core species representing clades I, III, and V of the Nematoda. These two resources allowed for direct assessment of intestinal genes/ proteins expressed by each of the phylogenetically diverse core species and transcriptomes from other tissues. A third essential omics resource was genome and predicted proteome information from each of the other nematode and outgroup species included

in the analysis. As a first step, orthologous protein groups were predicted from combined data for all species (10 nematodes and 5 outgroups) included in the analysis. Then, transcriptomic data from each core species were related to predicted proteome data from each and the orthologous protein groups derived from all 15 species. Intestinal protein families (IntFams) could then be inferred across orthologous protein families identified at any of numerous phylogenetic levels based on the expression from the core species. The relationship of the core species data to inferred intestinal proteomes from other nematode species is shown in **Figure 3**. The study based on intestinal expressed sequence tags identified 241 predicted IntFams from *H. contortus*, *A. suum*, and *C. elegans* (Yin et al., 2008) (based on 3,121 and 1,755 identified transcripts from *A. suum* and *H. contortus*, respectively), whereas the pan-Nematoda RNAseq study (Rosa et al., 2014) expanded the number of IntFams to 10,772 (based on complete gene sets from nematode species spanning the phylum). Phylogenetic assessment of IntFam gene births and deaths among the 10 species analyzed produced phylogenetic tree associations expected for trees of this species assemblage based on other quantitative criteria (Rosa et al., 2014).

The phylogenetic affinities and conservation (predicted orthologs) of the IntFams determined by this analysis were parsed into 14 categories, ranging from IntFams/functions that are conserved among all the core nematode species (**cIntFams**, 2,853 from intestinal transcriptome data) to the ones conserved in all 10 of the diverse nematode species investigated (**nem-cIntfams**, 1,863). One can rationalize the biological or practical importance for each of these groupings, including the taxonomically restricted ones. The most inclusive group for intestinal genes included orthologs conserved in all 10 nematodes and 5 outgroups (**uni-cIntFams**). Conservation of this kind might be viewed to preclude utility

FIGURE 3 | A detailed example of one bioinformatic workflow. Here, a list of conserved intestinal-expressed protein families is produced by intersecting several datasets. The workflow for producing the *Ascaris suum* genome assembly and annotation is shown but omitted for the other two core species.
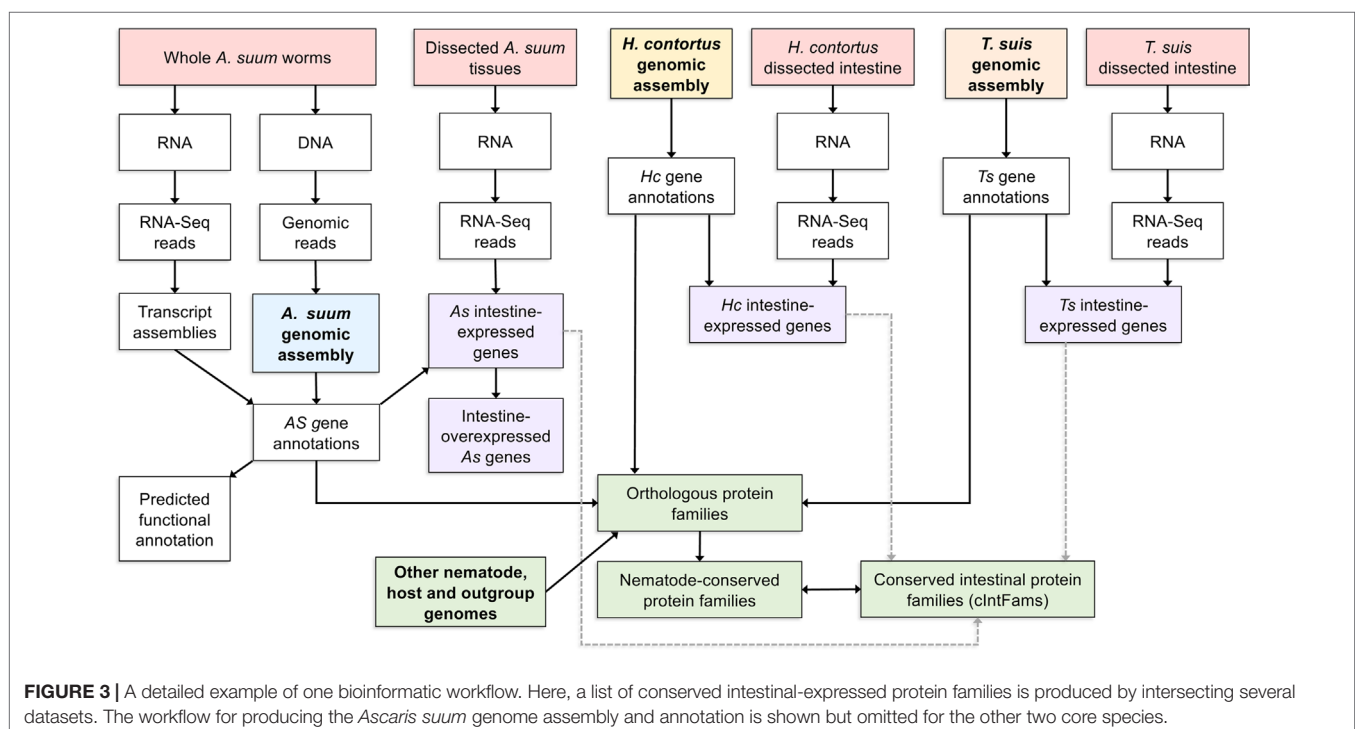
of proteins in therapeutic approaches. Nevertheless, insertion/ deletion (indel) analysis identified distinctions of potential practical significance between some nematode and mammalian members of these IntFams. Thus, the differences existed below the threshold utilized for parsing orthologous proteins. We can also expect for members of this group that amino acid substitutions may exist that are nemato-centric and have functional significance, such as in the case for many beta tubulins in nematodes (Hoti et al., 2003). Although this particular analysis was not conducted in a comprehensive manner, the database exists to do so. In some cases, IntFam members were detected based on transcripts from each of the core species (cIntFam) but not necessarily from genomes of all nematodes investigated. This level of conservation (clades I, III, and V) based on direct sampling of intestinal transcripts reflects broad conservation across many parasitic nematode species, although maybe not all. Some IntFams lacked representation in all core species (e.g., intestinal transcripts from all core species) but nevertheless had ortholog representation from genomes of all 10 nematode species evaluated. This situation might reflect false negatives (sampling artifact) from intestinal transcriptomes of individual core nematode species, or expression of some intestinal transcripts might be conditional and transient. Consequently, interpretation of the data warrants some caution, and the relationships described should be viewed as a beginning for developing hypotheses rather than a final determination of conservation.

Beyond information that can be gained on the molecular evolution of the intestine in adult parasitic nematodes, we envision multiple uses for this pan-Nematoda database of IntFams. From a basic omics perspective, the information renders less mysterious the specific intestinal proteins, protein isoforms, amino acid sequences, predicted functions, and diversity of those proteins that carry out essential functions of the intestine among nematode species. By organizing the data as predicted orthologs with specific proteins identified from each species investigated, researchers have a template to explore questions of individual interest and with some explicit gene and protein sequence information that can support experimental design efforts. As one example, protein isoforms from multimember families that are differentially expressed in the intestine or other tissues can be prospectively identified, by integrating the IntFam database and the comparative-tissue transcriptome database for *A. suum*, thus resolving substantial complexity ensconced in whole genome and predicted proteome data. Such resolution is critical when the biology of individual tissues is under investigation.
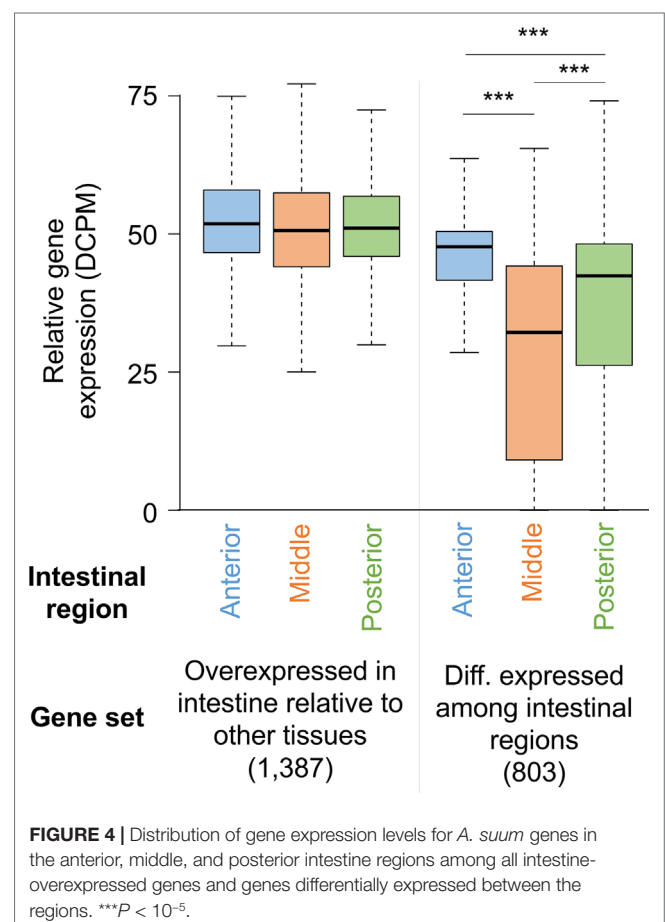
INCREASING RESOLUTION BY STUDYING INTESTINAL REGIONS AT A TRANSCRIPTOME AND PROTEOME LEVEL

Differential Expression of Genes Along the Length of the Intestine and Their Posttranscriptional Regulation

Another dimension of functional dissection involves differential expression of intestinal genes along the length of the intestine.

The anterior region of the *H. contortus* intestine is hypersensitive to benzimidazole treatments, by comparison to the posterior region (Jasmer et al., 2000). Consequently, biological differences among intestinal regions exist that appear to have practical applications. To study this in more depth, *A. suum* intestinal regions have been studied using RNAseq approach, resulting in resolving expression differences in a single, comprehensive effort, the results of which will enhance the use of this species in intestinal cell research and likely will have application to other nematode species.

Deep sequencing of transcripts obtained from contiguous anterior, middle, and posterior regions of the intestine from male and female *A. suum* worms expanded the set of previously known intestinal genes (Gao et al., 2016). No major differences were identified based on gender, with >80% intestinal genes expressed in both male and female. Genes expressed among the three regions were similar, with only 803 genes being differentially expressed. However, most of these (696/803) had higher expression levels in the anterior region as compared to the middle and posterior regions, supporting that the anterior intestinal region has certain functional distinctiveness as compared to the rest of the intestine (**Figure 4**). This difference was also indicated by assembling the genes based on different expression profiles along the intestinal axis. This less stringent analysis identified genes

FIGURE 4 | Distribution of gene expression levels for *A. suum* genes in the anterior, middle, and posterior intestine regions among all intestine-overexpressed genes and genes differentially expressed between the regions. ***$P < 10^{-5}$.

expressed at higher levels in the anterior region including those encoding certain hydrolases and those with functions related to signal transduction and membrane dynamics. Given that a vast majority of genes were expressed throughout the intestine, although most of these are expressed at relatively low levels, one way of delineating major intestinal functions is to analyze genes with relatively high overall intestinal expression. Based on a comparison with the expression levels of all genes, transcripts for 795 were considered to be highly expressed in the intestine. As expected, some of the functions encoded by these highly abundant intestinal transcripts were related to nutrient uptake and energy metabolism. Other enriched functions included protein synthesis and protein and lipid binding.

Samples from the same regions of the *A. suum* intestine were profiled for miRNAs. This provided another layer of possible differentiation among intestinal regions, i.e., regulators of intestinal gene expression. miRNAs were known to be expressed in gametogenesis and early developmental stages of *A. suum* (Wang et al., 2011). On the other hand, intestinal miRNAs have previously been identified in *C. elegans*, and these were shown to cause effective downregulation of mRNA expression (Kato et al., 2016). We leveraged the availability of samples from different intestinal segments to further expand the knowledge base of nematode miRNAs, their interactions with the transcriptome, and differential expression along the intestinal tract (Gao et al., 2016). As a result of this work, 277 miRNAs were identified in the *A. suum* intestine, and 11 of those were differentially expressed among intestinal regions. To integrate the information from mRNA and miRNA expression sets, potential targets of these miRNAs were predicted for the ~6,000 genes of the *A. suum* genome that had a 3'UTR sequence available and could be used for miRNA–target prediction. Out of these, 2,063 were predicted to be targeted by at least one intestinal miRNA. As is usually the case with predicted miRNA targets, the predicted miRNA-target network was dense with many-to-many targeting interactions commonly encountered (i.e., multiple targets associated with a given miRNA and vice versa). These predictions need validations, as does the existence of such a highly active regulatory network.

The large number of samples included in this study provided abundance data for both mRNAs and miRNAs that supported a novel statistical approach for predicting miRNA–target associations. Correlations between sample abundances of miRNAs and corresponding predicted mRNA targets identified 503 pairs categorized as most likely to be associated with each other (LAMPs: likely associated miRNA-mRNA pairs). In a similar vein, mean correlations of some miRNAs with corresponding target mRNAs were markedly higher than mean correlations of other miRNAs. These miRNAs were classified as the most likely influential miRNAs (LIMs). Interestingly, encoded proteins of many of the set of predicted targets for the 22 LIMs showed significant functional enrichment [GO (The Gene Ontology, 2019), InterPro (Mitchell et al., 2019), etc.], potentially indicating real biological functions under miRNA regulation. An analysis of miRNA mature and seed sequences identified some miRNAs and miRNA families whose intestinal expression, and hence, potential intestinal function is conserved between *A. suum* and *C. elegans*

or *Heligmosomoides bakeri* (Gao et al., 2016). The intestinal miRNA databases offer guidance to determine functions of these regulatory molecules in intestinal gene expression and functions of proteins encoded by target mRNAs.

Resolution of gene expression by transcriptomics along longitudinal regions of the intestine as done with *A. suum* is far more challenging for many other nematodes, parasitic or not. However, integrating information on differentially expressed genes and proteins across *A. suum* intestinal regions with the pan-Nematoda intestinal omics databases can lead to useful predictions on intestinal expression patterns of orthologous genes and proteins from other nematode species.

Intestinal Functions Detected at a Protein Level

Traditional research on intestinal antigens targeted in vaccine research has identified a modest number of AIM proteins, including many proteases (Andrews et al., 1995; Jones and Hotez, 2002; Smith et al., 2003a; Smith et al., 2003b; Williamson et al., 2003a; Williamson et al., 2003b; Loukas et al., 2004; Williamson et al., 2004; Jasmer et al., 2007; Knox, 2011). Factors that hindered more comprehensive identification of AIM functions include the relative narrow focus of those investigations, available methods, and limitations posed by the nematodes investigated. In contrast, more recent studies took advantage of the large size of *A. suum*, which when coupled with extensive databases of intestinal genes/predicted proteins, supported a proteomic approach that greatly increased knowledge of proteins and predicted functions on the AIM and other cellular compartments of the intestine. Cannulation of the adult *A. suum* intestine (**Figure 5**), below the pharynx, with ordinary blunt-ended hypodermic needles (Rosa et al., 2015) facilitated the perfusion of the intestine with, in this case, phosphate-buffered saline (PBS) to directly collect lumen content and directly identify proteins in this perfusate by mass spectrometry (**Figure 6**). Intestinal antigen research in *H. contortus* (Jasmer et al., 2007) indicated that many of the intestinal proteases are associated with the AIM (Williamson et al., 2003b), but because those proteins had predicted signal peptides and most often lacked evidence of transmembrane regions, it was suggested that they function as peripheral membrane proteins. Thus, *A. suum* intestinal lumen was also perfused with 4 M urea (4 MU) to solubilize peripheral membrane proteins. Additionally, the glycans on many of the *H. contortus* AIM proteases have figured centrally in methods used to isolate, identify, and functionally characterize these proteins (Jasmer et al., 1993; Smith et al., 1993; Smith et al., 1994; Rosa et al., 2015). Similarly, the lectin concanavalin A predominantly bound to the AIM in *A. suum* intestinal tissue and was used to isolate and characterize apparent AIM glycoproteins by lectin affinity chromatography. These analyses of intestinal perfusates, lectin binding fractions, additional intestinal membrane fractions, and whole intestinal lysates by mass spectrometry have identify over 1,000 intestinal proteins that were assigned to cellular/tissue compartments based on methods of generating each fraction, and predicted physical properties including, charge, signal peptides, and transmembrane regions.
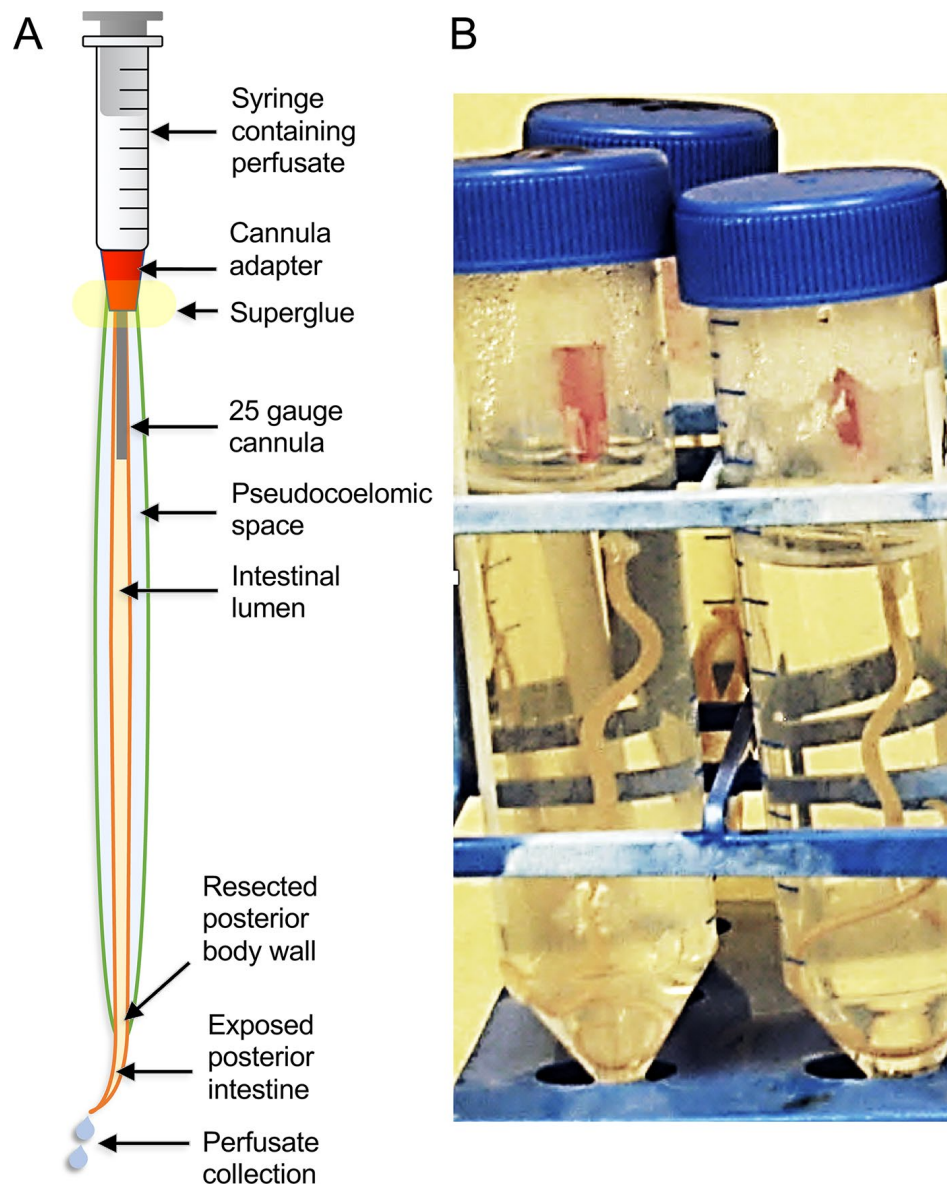
FIGURE 5 | The *A. suum* intestinal cannulation and perfusion system. Shown are **(A)** a diagram description of the system and **(B)** a picture of the actual worms set up in the system, and contained inside of plastic test tubes.

This approach identified numerous proteins located in the *A. suum* lumen or on the AIM along with predicted functions of those proteins (Rosa et al., 2015). Experimental evidence was also gained on a range of proteases and related activities that were detected in perfusates of the intestinal lumen (Jasmer et al., 2015). About 157 distinct proteins were obtained in the PBS and 4 MU perfusates. Two major groups of proteins attracted the greatest interest in these fractions, digestive hydrolases and channel/transport proteins, which are expected to contribute to ion and nutrient transport. The hydrolases included proteases (28 in 5 classes of proteases) with metallopeptidases being best represented, including 16 different metallopeptidases. Aspartic proteases and serine carboxypeptidases were the next best

represented with five and four proteins, respectively, followed by cysteine and threonine proteases representing the remainder. The relative abundance and diversity of apparent proteases identified in these fractions answered questions regarding the comparative need for protein digestion in *A. suum* which lives in the host small intestine where host protein digestion is nearing completion. Apparently, *A. suum* has substantial need for digestion of proteins consumed in this host location. O-Glycosyl hydrolases comprised a second group of hydrolases, which likely contribute to saccharidase activity previously investigated in *A. suum* (Gentner and Castro, 1974), and some of these hydrolases were identified in previous investigations on *A. suum* larval stages and intestinal tissue (Wang et al., 2013a). A third group of hydrolases
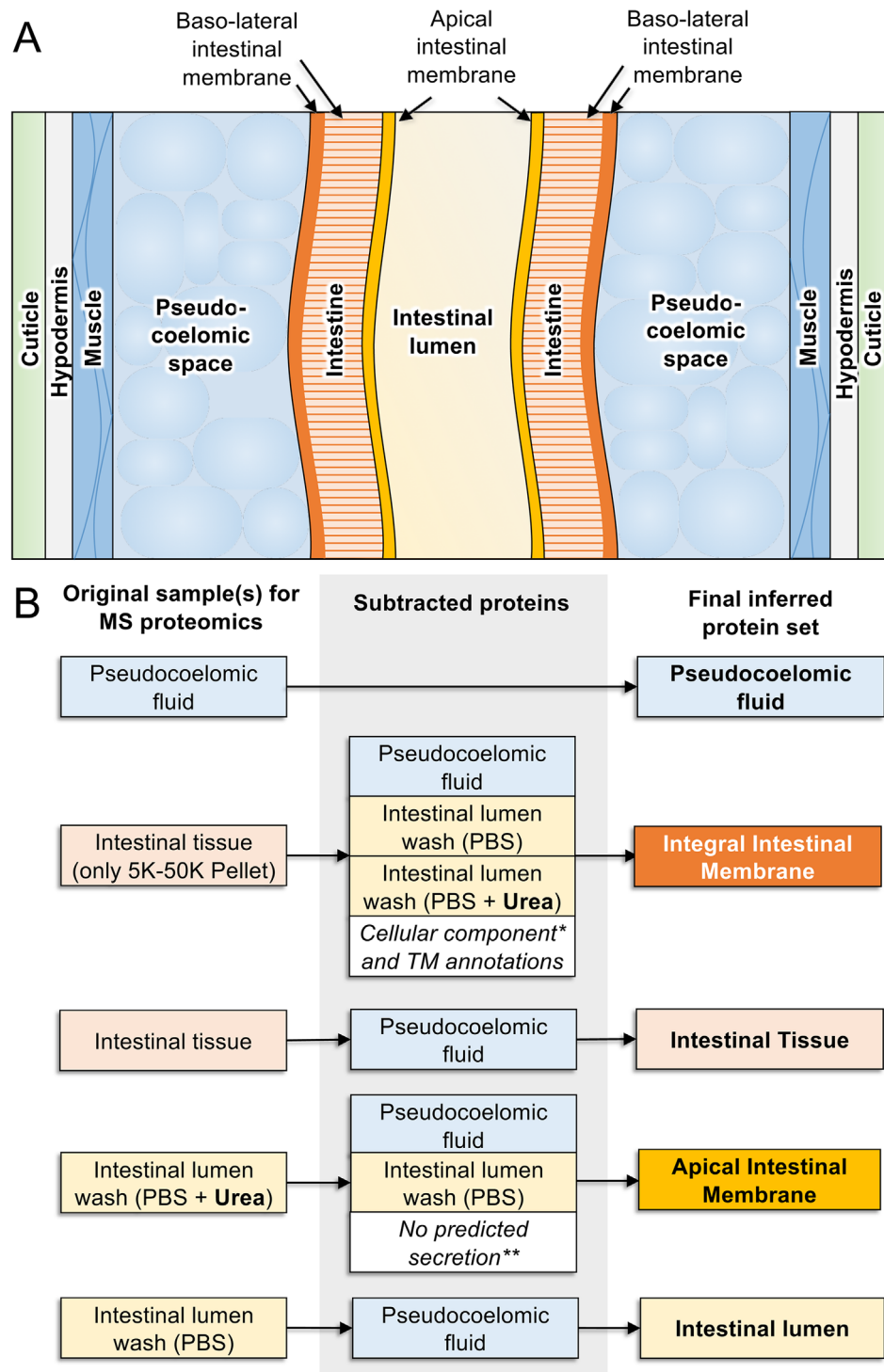
FIGURE 6 | Proteomics-based inference of *A. suum* intestinal proteins in different compartments. **(A)** Anatomy of model intestinal nematode species *Ascaris suum* (transverse section). **(B)** Protein sets detected by MS/MS proteomics from samples harvested from adult *A. suum* worms (left) have other protein sets strategically removed (center) to deduce final protein sets in different intestinal compartments (right). "Integral intestinal membrane" proteins are not labelled as "basal" because they may include some proteins from the apical intestinal membrane as well. *Proteins annotated with "cellular compartment" Gene Ontology terms for endoplasmic reticulum, mitochondria, Golgi apparatus, and nucleus were removed to reduce contamination from proteins embedded in these organelles rather than the external cellular membrane. **Only proteins annotated with predicted classical or nonclassical secretion signals were included since these are better candidates for proteins that are transported to the membrane.

included several lipases, thus rounding out a basic set of digestive enzymes in the lumen of *A. suum* intestine. The study detected many other predicted functions that fall outside these groups. These collective data can help formulate ideas to target digestive processes in anthelmintic strategies against *A. suum* and possibly other parasitic nematodes.

The detected channel/transport proteins that mediate transport of ions and nutrients across intestinal membranes are of interest because interference with these functions has potential for anthelmintic approaches, and new information about them should facilitate development of these approaches. Information in this context was largely derived from the 4 MU perfusates and the membrane preparation, which included many proteins detected in the 4 MU perfusate as a subset. Over 100 proteins were classified as channel/transport proteins, with predicted transmembrane (integral membrane proteins) or other annotation supporting this assignment, and detection in membrane enriched fractions added further support. However, many of the channel/transport proteins were detected only in the membrane fraction, which does not resolve apical, basolateral, or intracellular membrane localization in intact intestine, although an attempt was made to exclude proteins associated with cytoplasmic organelles. Distinguishing membrane location is important particularly if antibody-mediated inhibition is a goal in vaccine research, as this would likely require localization to the AIM. Accessibility by localization to the AIM may also be an important factor for pharmacotherapy, which becomes much more feasible with knowledge of potential targets that are accessible on the AIM. An example here is the apical localization of the receptor for CRY protein toxins, the absence of which on the AIM is likely to render CRY proteins less effective (Griffitts et al., 2003; Williamson et al., 2003b). Nevertheless, for a large number of predicted integral membrane proteins identified, channel/transport proteins and others, the actual membrane where they reside remains to be determined. Alternatively, some of these membrane proteins were also, or exclusively, detected in the 4 MU fraction, which provided evidence for an AIM localization, although not to the exclusion of the basolateral intestinal membrane (BIM), also. One example highlighted by this research involves subunits of the vacuolar ATPase V1 domain. Seven of the eight subunits were detected in the 4 MU fractions, which likely reflects solubilization by 4 MU (Rosa et al., 2015). The results provide evidence that similar to *C. elegans* (Allman et al., 2009), V-ATPases occupy the AIM and contribute to H+ transport across this surface in *A. suum*. This kind of evidence was obtained for 57 other channel/transport proteins, nevertheless leaving the majority of the predicted integral membrane proteins with little direct evidence for the membrane (s) on which they function.

Examining the sequence conservation of the orthologous proteins detected in different cellular compartments across species spanning the phylum Nematoda revealed significant variability. Most variable were the proteins detected in the intestinal lumen (IL), and integral intestinal membrane proteins were the most conserved among nematodes. Concurrently, proteins homologous to most of the *A. suum* IL proteins were detected in other nematode species, and because hydrolytic enzymes constitute a substantial subset of IL proteins, the result may reflect diversification from common ancestral digestive enzymes among species. Although neutral evolution could account for diversification in this background, directional evolution might be expected if adaptations are required of digestive enzymes for a given species to best exploit nutrients presented by different host niches. Hemoglobinases from hookworms may exemplify this idea in that human hemoglobin and serum proteins were digested more efficiently than orthologous proteins from dogs by the hemoglobinase Na-APR-2 from the hookworm of humans, *Necator americanus* (Williamson et al., 2003a). Despite the lower level conservation for individual protease proteins, major classes of IL proteases (aspartic, cysteine, metallo-, and serine) appear to be conserved among many nematode species and appear to retain broad functional characteristics related to inhibitors of the protease classes, which may have pharmacotherapeutic applications. The omics data utilized to identify probable IL proteins from other nematodes may have application toward identification of IL proteins from other nematode pathogens. It is important to note that many of the predicted proteases that colocalized to the PBS and 4 MU fractions were classified as IL proteins, even though they may function as peripheral membrane proteins that then are released into the lumen.

The proteomics analysis summarized here significantly clarifies functions that might be sited in the various intestinal cell and tissue compartments, and available evidence can be consulted for prospective development of research hypotheses stimulated by information in this resource. As one example, much of the past vaccine research has focused on proteases and their inhibition by host neutralizing antibodies ingested by the parasite. Alternatively, channel/transport proteins sited on the AIM have at least as much attraction as hydrolases for antibody-mediated inhibition of nutrient acquisition and other homeostatic functions (such as ion transport). The numerous specific examples of prospective AIM channel/transport proteins can now be prioritized for investigation in this direction. Additional membrane proteins were identified, some of which likely function on the AIM and can be investigated individually. However, and despite the progress reported, with only about 1,000 intestinal proteins identified by mass spectrometry, the advances are modest by comparison to what can be achieved with *A. suum* in a more concerted proteomic analyses of nematode intestinal cell compartments. Not many nematode species can support the methods and tissue demands needed to gain the insight offered on this topic by the *A. suum* intestinal system. Given that the tools exist to place resulting information into a pan-Nematoda context, this area should be exploited to the full extent that the *A. suum* system can support.

We provide here an example of one of the many ways the generated multiomics database can be interrogated. Published reports provide supplementary tables in which the data have been integrated into a database that includes (for every documented *A. suum* gene) functional annotation, pan-Nematoda phylogeny, RNAseq expression and differential expression, and proteomic data spanning all of the available datasets (**Table 2**) from the publications discussed in this review. This resource allows for straightforward and convenient cross-study dataset comparisons

and the prioritization of genes of interest to suit the goals of new studies.

In addition to our ongoing *A. suum* intestinal omics research, other nematode intestinal omics datasets are available in the literature including a transcriptomics study of the *Ancylostoma ceylanicum* intestine, a proteomic study comparing the intestine, body wall, and reproductive tissues of *B. malayi*, and a proteomic study of larval-stage *A. suum* excretory–secretory (ES) products that provides more information about potential intestine-produced products during the larval stage (Wang et al., 2013a). As we have performed for our projects, future research will benefit from intersecting the results from these valuable intestinal omics resources, to provide more information about genes of interest and better insights into conserved and specific intestinal functions.

ADVANCES ON METHODS TO EXPERIMENTALLY MANIPULATE *A. SUUM* INTESTINAL CELLS

The Intestinal Cannulation and Perfusion System

As with many aspects of helminth research, the small size of many nematodes, limited access to live parasite tissues of interest, and poor *in vitro* survival of the life cycle stages of interest all present challenges for investigations on intestinal cells. Again, *A. suum* presents distinct advantages relative to obtaining abundant biological materials for biochemical analyses. Its large size has facilitated improvements to experimentally investigate intestinal cell functions, particularly as relates to cannulation and perfusion of the intestine (**Figure 5**). This technique supports delivery of controlled treatments (amount, timing) into the lumen for experimental manipulation of the AIM and other intestinal cell functions. Delivery *via* perfusion leads to the immediate achievement of treatment concentrations in the lumen. In contrast, delivery by feeding is variable and difficult to quantify or is unachievable for many parasitic species. Maintenance of the cannulated intestine *in situ* where it is bathed by pseudocoelomic fluid leads to expectations for faithful replication of *in vivo* functions in this setup, particularly over short-term experiments. Cannulated *A. suum* can be maintained in culture media for at least 4 days after which time worms are still motile, although stable maintenance of intestinal cell functions during this time period requires verification. Experimental treatments can also be injected into the pseudocoelom for delivery across the BIM. The advantage of BIM delivery is that pseudocoelomic injections can be done with intact worms, which can then be maintained for several days in *in vitro* culture. Pseudocoelomic delivery of bacteria was shown to induce expression of intestinal transcripts encoding *A. suum* antibacterial factors (ASABF) (Pillai et al., 2003) as one example. However, in many cases, transport across the AIM is not expected to be replicated by the BIM, and transporters may have different polarities if found in both the AIM and BIM. Because BIM delivery of experimental treatments to intestinal cells is not an obvious alternative to dissect AIM functions,

simple methods like cannulation and perfusion of the intestinal lumen represent valuable capabilities to investigate functions peculiar to the AIM. It is also likely that the *A. suum* system will support experimental dissection of functions on the BIM, about which even less is known. Nevertheless, double-stranded RNA (dsRNA) treatments delivered by both routes causes knockdown of target intestinal cell transcripts, as discussed below. Hence, the two different routes of delivery may have complementary characteristics for at least some treatments, as described in the next paragraph.

Manipulation of Intestinal Cell Transcripts by Perfusion of dsRNA

The ability of the intestinal perfusion method to support experimental manipulation of intestinal cell functions has been tested using dsRNA as the treatment for several *A. suum* intestinal genes (Rosa et al., 2017). The genes used in this study were selected based on a number of criteria, including evidence of results with dsRNA from orthologous genes in heterologous species. A series of experiments was conducted in which dsRNA constructs were tested for each of five genes. Based on the results, dsRNA was selected for one of these genes (GS_08011) to assess treatment variables, including dsRNA construct (intact long or dsRNA fragmented by ribonuclease III), amount, and comparative effectiveness of delivery across the AIM or BIM. These studies showed that while knockdown of transcripts was impressive for some genes (up to 99.7% expression reduction), there was also some variability among genes, which is not unusual for *A. suum* or other parasitic nematodes (Selkirk et al., 2012). Perfused quantities down to 0.5 µg per worm caused effective knockdown of transcripts, and knockdown was achieved to approximately the same level in anterior and posterior halves of the intestine. Likewise, knockdown of intestinal transcripts was achieved to approximately the same level for dsRNA perfused into the intestinal lumen or injected into the pseudocoelom (the effect of which was previously investigated for pseudocoelomic delivery of dsRNA) (McCoy et al., 2015). These effects were achieved within a 24-h treatment period.

Assessment of Off-Target dsRNA Effects

It is important to evaluate off-target dsRNA effects by RNAseq and challenges they might present for interpretation of experimental results even when effectiveness of dsRNA delivery and knockdown of intestinal cell transcripts is established. The *A. suum* dsRNA perfusion study (Rosa et al., 2017) quantified global transcript effects for treatments of dsRNA constructs from two different genes, then control decapitated worms at time 0 or cannulated for 24 h, and anterior and posterior regions of worms from each treatment group. By comparison to other species/model systems, only modest effects were observed for up- or downregulation of off-target genes related to dsRNA treatments. For instance, transcripts for only eight and four intestinal genes showed significant up- or downregulation, respectively, by both dsRNA gene constructs in both anterior and posterior intestinal regions (Rosa et al., 2017).

Additional, but still modest, effects were observed related to individual dsRNA gene treatments and then anterior and posterior intestinal regions. Overall, off-target effects will require attention in relation to functional studies, perhaps in the form of using multiple distinct dsRNA constructs to rule out construct-specific effects and incorporating results from complementary approaches for functional analyses.

Summary of Progress on the *A. suum* Intestinal Research Model

Attempts to manipulate *A. suum* intestinal cells using the perfusion system have been successful. Overall, the published results indicate that the system provides a relatively facile method to manipulate intestinal cell functions, in this case knockdown of target transcripts by dsRNA. Although concomitant knockdown of protein will be an important follow-up to assess the utility of the dsRNA approach to determine protein functions, the research demonstrated the ability of the system to support exogenous manipulation of intestinal cells. The intestinal perfusion system should enable many questions to be investigated regarding functions that reside in the lumen, on the AIM, and in intestinal cells, thus adding motivation to gain a more complete picture of compartments occupied by proteins expressed by intestinal cells. In context of omics resources for the *A. suum* intestinal model coupled with pan-Nematoda IntFam resources, *A. suum* offers valuable attributes to a cooperative multispecies model aimed to derive essential functions of intestinal cells common to many parasitic nematodes. It remains to be determined how the other two core species (*H. contortus* and *T. suis*) will add to the cooperative model. Concomitantly, one can now predict intestinal genes in other parasitic nematodes, and *Strongyloides* spp. provide parasite examples where transgenic methods can be used to test functional predictions *in vivo* (Shao et al., 2017). Pan-Nematoda IntFams also provide a connection that can guide the use of *C. elegans* to investigate intestinal functions with broad application to nematode pathogens.

THE FUTURE OF PARASITIC NEMATODE INTESTINE RESEARCH

Further Development of the *A. suum* Experimental System

The *A. suum* experimental system is well recognized as an important model for parasite research. Creative approaches by many investigators have capitalized on the advantages that this large parasite offers; yet, the potential of the model is far from fully realized. Development of the model for research on intestinal cell biology is at an early stage, with progress driving desire for much more detail, and extensive intestinal omics data will facilitate derivation of many hypotheses. For instance, while longitudinal queries by RNAseq delineate rather gross assessment of differentiation along the intestinal tract, functional nuances among individual intestinal cells interrogated by single cell sequencing (Hwang et al., 2018) remain an intriguing goal to achieve. Major interests of ours include AIM functions and intracellular processes that mediate secretion and renewal of functions that are sited on the AIM. Additionally,

more recent findings have implicated exosomal vesicles containing miRNAs derived from the *H. polygyrus* intestine (presumably secreted from the AIM) in modifying host immune responses (Buck et al., 2014; Coakley et al., 2017). Improving the knowledge base and experimental methods for the nematode intestine may contribute to research on the worm side of the interaction. In these contexts, pan-Nematoda omics data coupled with the cannulation/perfusion system offer substantial capabilities to investigate hypotheses related to AIM biology, and efforts are needed to identify the best ways to maintain and manipulate intestinal cells outside of the parasite. Although proteomics data provide valuable information on AIM constituents, this resource can and should be expanded to more comprehensively document functions on the AIM. The adult intestinal cell system will support many experiments; however, the effort and cost to obtain adult *A. suum* are not without drawbacks. Larval stages of *A. suum* (lung stage larvae or early intestinal stages) can provide a less expensive and more ready source of parasite material and are likely to support investigation of some hypotheses on intestinal cells while extending application to additional life cycle stages. Larval stages have generated some useful findings relevant to intestinal cell biology (Wang et al., 2013a). Thus, effort is needed to establish capacity to utilize larval stages for experimental purposes directed at intestinal cell functions. The list can go on, but the areas mentioned are the ones that have priority for us.

Drug Targets and Therapies

Omics data for intestinal cells and compartments, such as the AIM, provide enormous potential when integrated with drug and target databases to identify prospective parasite targets for existing inhibitors. Areas that are ripe for investigation included enzymes and membrane transporters localized to the AIM. Alternatively, an approach we are following involves predictions of essential cellular pathways and components, which preferentially are pan-Nematoda conserved. One example here involves pathways that transport secretory vesicles containing digestive enzymes and other cargo to the AIM. The goal is to more broadly clarify parasite components that are involved in the exocytic process and identify inhibitors that might replicate irreparable intestinal damage caused by benzimidazole anthelmintics. Whether or not new anthelmintics result, experimentation of this kind is likely to produce new information on basic intestinal cell biology that has potential for pan-Nematoda applications. Further, although the research is intestino-centric, results may have applications to other tissues and cells. Using secretion as an example, secretions in intestinal cells and neurons likely have similar and perhaps identical cellular components, although the consequences of inhibition may be idiosyncratic to a given tissue (tissue degeneration, or disruption of neurotransmission, respectively). There is evidence in *C. elegans* that some proteins that mediate/regulate exocytosis (e.g., syntaxin) have similar functions in each tissue type. Thus, learning more about intestinal cell secretion and inhibitors may have application to other parasite tissues, including the nervous system.

Vaccine Targets and Immune Control

Development of subunit vaccines against parasitic helminths (with some exception of tapeworms) has proven an elusive

goal. Although not without caveats, given the remarkable levels of protection using intestinal antigens against blood-feeding nematodes, progress summarized has identified multiple categories of proteins with potential application to immune control. In addition to likely digestive enzymes, numerous predicted integral membrane and transport proteins were identified that are likely to be localized to the AIM, and host antibody consumed by the parasite would have access to them. Many of these proteins have predicted orthologs in other nematode species, which creates opportunities for research in those species. This dataset, and ideally in the future a larger dataset, identifies numerous functions that can be investigated with antibody-mediated inhibition of those functions. Recognizing that the best immunization results have been achieved thus far in context of blood-feeding parasites, the approach may also have application to somatically migrating stages of parasites such as *A. suum*, which will be in contact with antibodies found in serum. Transcripts for many adult intestinal genes have already been identified, inclusive of those encoding AIM-localized proteins. These genes are expressed in migratory larvae as well, although it remains unclear how extensively these larvae rely on feeding during their migration to the intestine. In any case, the omics data presented stimulate many lines of thought on possible applications of this information to future vaccine development related studies.

AUTHOR CONTRIBUTIONS

DJP, RT, and MM all contributed to the writing and all approved the content of this review article.

FUNDING

REFERENCES

Allman, E., Johnson, D., and Nehrke, K. (2009). Loss of the apical V-ATPase a-subunit VHA-6 prevents acidification of the intestinal lumen during a rhythmic behavior in C. elegans. *Am. J. Physiol. Cell Physiol.* 297 (5), 9. doi: 10.1152/ajpcell.00284.2009

Andrews, S. J., Hole, N. J., Munn, E. A., and Rolph, T. P. (1995). Vaccination of sheep against haemonchosis with H11, a gut membrane-derived protective antigen from the adult parasite: prevention of the periparturient rise and colostral transfer of protective immunity. *Int. J. Parasitol.* 25 (7), 839–846. doi: 10.1016/0020-7519(94)00221-9

Anisimov, A. P., Tokmakova, N. P., and Usheva, L. N. (1975). Proliferation and growth of intestinal epithelium in Ascaris suum (Nematoda) during postnatal ontogeny. Communication III. Mitotic anomalies and changes in ploidy and nucleus size. *Sov. J. Dev. Biol.* 5 (1), 37–44.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. the gene ontology consortium. *Nat. Genet.* 25 (1), 25–29. doi: 10.1038/75556

Atkinson, C., Newsam, R. J., and Gull, K. (1980). Influence of the antimicrotubule agent, mebendazole, on the secretory activity of intestinal cells of Ascaridia galli. *Protoplasma* 105 (1-2), 69–76. doi: 10.1007/BF01279850

Ballesteros, C., Geary, J. F., Mackenzie, C. D., and Geary, T. G. (2018). Characterization of Divalent Metal Transporter 1 (DMT1) in Brugia malayi suggests an intestinal-associated pathway for iron absorption. *Int. J. Parasitol. Drugs Drug Resist.* 8 (2), 341–349. doi: 10.1016/j.ijpddr.2018.06.003

Basavaraju, S. V., Zhan, B., Kennedy, M. W., Liu, Y., Hawdon, J., and Hotez, P. J. (2003). Ac-FAR-1, a 20 kDa fatty acid- and retinol-binding protein secreted by adult Ancylostoma caninum hookworms: gene transcription pattern, ligand binding properties and structural characterisation. *Mol. Biochem. Parasitol.* 126 (1), 63–71. doi: 10.1016/S0166-6851(02)00253-0

Bassetto, C. C., Almeida, F. A., Newlands, G. F. J., Smith, W. D., Castilhos, A. M., Fernandes, S., et al. (2018). Trials with the Haemonchus vaccine, Barbervax((R)), in ewes and lambs in a tropical environment: nutrient supplementation improves protection in periparturient ewes. *Vet Parasitol.* 264, 52–57. doi: 10.1016/j.vetpar.2018.11.006

Borgers, M., and De Nollin, S. (1975). Ultrastructural changes in Ascaris suum intestine after mebendazole treatment *in vivo*. *J. Parasitol.* 61 (1), 110–122. doi: 10.2307/3279120

Borgers, M., De Nollin, S., De Brabander, M., and Thienpont, D. (1975). Influence of the anthelmintic mebendazole on microtubules and intracellular organelle movement in nematode intestinal cells. *Am. J. Vet. Res.* 36 (08), 1153–1166.

Bradley, J. E., Nirmalan, N., Kläger, S. L., Faulkner, H., and Kennedy, M. W. (2001). River blindness: a role for parasite retinoid-binding proteins in the generation of pathology? *Trends Parasitol.* 17 (10), 471–475. doi: 10.1016/S1471-4922(01)02036-0

Buck, A. H., Coakley, G., Simbari, F., McSorley, H. J., Quintana, J. F., Le Bihan, T., et al. (2014). Exosomes secreted by nematode parasites transfer small RNAs to mammalian cells and modulate innate immunity. *Nat. Commun.* 5, 5488. doi: 10.1038/ncomms6488

Cappello, M., Bungiro, R. D., Harrison, L. M., Bischof, L. J., Griffitts, J. S., Barrows, B. D., et al. (2006). A purified Bacillus thuringiensis crystal protein with therapeutic activity against the hookworm parasite Ancylostoma ceylanicum.. *Proc. Natl. Acad. Sci. U.S.A.* 103 (41), 15154–15159. doi: 10.1073/pnas.0607002103

Coakley, G., McCaskill, J. L., Borger, J. G., Simbari, F., Robertson, E., Millar, M., et al. (2017). Extracellular vesicles from a helminth parasite suppress macrophage activation and constitute an effective vaccine for protective immunity. *Cell Rep.* 19 (8), 1545–1557. doi: 10.1016/j.celrep.2017.05.001

Coghlan, A., Tyagi, R., Cotton, J. A., Holroyd, N., Rosa, B. A., Tsai, I. J., et al. (2019). Comparative genomics of the major parasitic worms. *Nat. Genetics.* 51 (1), 163–174. doi: 10.1038/s41588-018-0262-1

Comley, J. C. (1980). Ultrastructure of the intestinal cells of Aspiculuris tetraptera after in vivo treatment of mice with mebendazole and thiabendazole. *Int. J. Parasitol.* 10 (2), 143–150. doi: 10.1016/0020-7519(80)90026-0

Cotton, J. A., Bennuru, S., Grote, A., Harsha, B., Tracey, A., Beech, R., et al. (2016). The genome of Onchocerca volvulus, agent of river blindness. *Nat. Microbiol.* 2, 16216. doi: 10.1038/nmicrobiol.2016.216

Eisenback, J. D. (1985). "Detailed morphology and anatomy of second-stage juveniles, males and females of the genus Meloidogyne (root-knot nematodes)," in *In An Advanced Treatise on Meloidogyne: Biology and Control*. Eds. J. N. Sasser and C. C. Carter. (Raleigh: North Carolina State University Graphics).

Fernandez, A. G., Gunsalus, K. C., Huang, J., Chuang, L. S., Ying, N., Liang, H. L., et al. (2005). New genes with roles in the C. elegans embryo revealed using RNAi of ovary-enriched ORFeome clones. *Genome Res.* 15 (2), 250–259. doi: 10.1101/gr.3194805

Gao, X., Tyagi, R., Magrini, V., Ly, A., Jasmer, D. P., and Mitreva, M. (2016). Compartmentalization of functions and predicted miRNA regulation among contiguous regions of the Nematode Intestine. *RNA Biol.* 14 (10), 1335–1352 doi: 10.1080/15476286.2016.1166333

Gentner, H. W., and Castro, G. A. (1974). Origin of intestinal disaccharidases of Ascaris suum. *Exp. Parasitol.* 35 (1), 125–131. doi: 10.1016/0014-4894(74)90015-0

Griffitts, J. S., Huffman, D. L., Whitacre, J. L., Barrows, B. D., Marroquin, L. D., Müller, R. et al., (2003). Resistance to a bacterial toxin is mediated by removal of a conserved glycosylation pathway required for toxin–host interactions. *J Biol Chem.* 278, 46, 45594–45602. doi: 10.1074/jbc.M308142200

Hanser, E., Mehlhorn, H., Hoeben, D., and Vlaminck, K. (2003). *In vitro* studies on the effects of flubendazole against Toxocara canis and Ascaris suum. *Parasitol. Res.* 89 (1), 63–74. doi: 10.1007/s00436-002-0668-6

Hedgecock, E. M., and White, J. G. (1985). Polyploid tissues in the nematode Caenorhabditis elegans. *Dev. Biol.* 107 (1), 128–133. doi: 10.1016/0012-1606(85)90381-1

Hotez, P. J., Diemert, D., Bacon, K. M., Beaumier, C., Bethony, J. M., Bottazzi, M. E., et al. (2013). The human hookworm vaccine. *Vaccine* 31 Suppl 2, B227–232. doi: 10.1016/j.vaccine.2012.11.034

Hoti, S. L., Subramaniyan, K., and Das, P. K. (2003). Detection of codon for amino acid 200 in isotype 1 beta-tubulin gene of Wuchereria bancrofti isolates, implicated in resistance to benzimidazoles in other nematodes. *Acta Trop.* 88 (1), 77–81. doi: 10.1016/S0001-706X(03)00159-1

Howe, K. L., Bolt, B. J., Shafie, M., Kersey, P., and Berriman, M. (2017). WormBase ParaSite—a comprehensive resource for helminth genomics. *Mol. Biochem. Parasitol.* 215, 2–10. doi: 10.1016/j.molbiopara.2016.11.005

Howells, R. E., and Chen, S. N. (1981). Brugia pahangi: feeding and nutrient uptake *in vitro* and *in vivo*. *Exp. Parasitol.* 51 (1), 42–58. doi: 10.1016/0014-4894(81)90041-2

Howells, R. E., Mendis, A. M., and Bray, P. G. (1983). The mode of action of suramin on the filarial worm *Brugia pahangi*. *Parasitology* 87 (Pt 1), 29–48. doi: 10.1017/S0031182000052392

Hu, Y., and Aroian, R. V. (2012). Bacterial pore-forming proteins as anthelmintics. *Invert. Neurosci.* 12 (1), 37–41. doi: 10.1007/s10158-012-0135-8

Huffman, D. L., Bischof, L. J., Griffitts, J. S., and Aroian, R. V. (2004). Pore worms: using Caenorhabditis elegans to study how bacterial toxins interact with their target host. *Int. J. Med. Microbiol.* 293 (7–8), 599–607. doi: 10.1078/1438-4221-00303

Hunter, S., Jones, P., Mitchell, A., Apweiler, R., Attwood, T. K., Bateman, A., et al. (2012). InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res.* 40, D306–312. doi: 10.1093/nar/gkr948

Hwang, B., Lee, J. H., and Bang, D. (2018). Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp. Mol. Med.* 50 (8), 96. doi: 10.1038/s12276-018-0071-8

Jasmer, D. P., Lahmers, K. K., and Brown, W. C. (2007). Haemonchus contortus intestine: a prominent source of mucosal antigens. *Parasite Immunol.* 29 (3), 139–151. doi: 10.1111/j.1365-3024.2006.00928.x

Jasmer, D. P., Mitreva, M. D., and McCarter, J. P. (2004). mRNA sequences for Haemonchus contortus intestinal cathepsin B-like cysteine proteases display an extreme in abundance and diversity compared with other adult mammalian parasitic nematodes. *Mol. Biochem. Parasitol.* 137 (2), 297–305. doi: 10.1016/j.molbiopara.2004.06.010

Jasmer, D. P., Perryman, L. E., Conder, G. A., Crow, S., and McGuire, T. (1993). Protective immunity to Haemonchus contortus induced by immunoaffinity isolated antigens that share a phylogenetically conserved carbohydrate gut surface epitope. *J. Immunol.* 151 (10), 5450–5460.

Jasmer, D. P., Perryman, L. E., and McGuire, T. C. (1996). Haemonchus contortus GA1 antigens: related, phospholipase C-sensitive, apical gut membrane proteins encoded as a polyprotein and released from the nematode during infection. *Proc. Natl. Acad. Sci. U.S.A.* 93 (16), 8642–8647. doi: 10.1073/pnas.93.16.8642

Jasmer, D. P., Rosa, B. A., and Mitreva, M. (2015). Peptidases compartmentalized to the Ascaris suum intestinal lumen and apical intestinal membrane. *PLoS Negl. Trop. Dis.* 9 (1), e3375. doi: 10.1371/journal.pntd.0003375

Jasmer, D. P., Yao, C., Rehman, A., and Johnson, S. (2000). Multiple lethal effects induced by a benzimidazole anthelmintic in the anterior intestine of the nematode Haemonchus contortus. *Mol. Biochem. Parasitol.* 105 (1), 81–90. doi: 10.1016/S0166-6851(99)00169-3

Jex, A. R., Liu, S., Li, B., Young, N. D., Hall, R. S., Li, Y., et al. (2011). Ascaris suum draft genome. *Nature* 479 (7374), 529–533. doi: 10.1038/nature10553

Jones, B. F., and Hotez, P. J. (2002). Molecular cloning and characterization of Ac-mep-1, a developmentally regulated gut luminal metalloendopeptidase from adult Ancylostoma caninum hookworms. *Mol. Biochem. Parasitol.* 119 (1), 107–116. doi: 10.1016/S0166-6851(01)00409-1

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30 (9), 1236–1240. doi: 10.1093/bioinformatics/btu031

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457–D462. doi: 10.1093/nar/gkv1070

Kaplan, R. M. (2004). Drug resistance in nematodes of veterinary importance: a status report. *Trends Parasitol.* 20 (10), 477–481. doi: 10.1016/j.pt.2004.08.001

Kato, M., Kashem, M. A., and Cheng, C. (2016). An intestinal microRNA modulates the homeostatic adaptation to chronic oxidative stress in C. elegans.. *Aging (Albany NY)* 8 (9), 1979–2005. doi: 10.18632/aging.101029

Knox, D. (2011). Proteases in blood-feeding nematodes and their potential as vaccine candidates. *Adv. Exp. Med. Biol.* 712, 155–176. doi: 10.1007/978-1-4419-8414-2_10

Knox, D. P., Redmond, D. L., Newlands, G. F., Skuce, P. J., Pettit, D., and Smith, W. D. (2003). The nature and prospects for gut membrane proteins as vaccine candidates for Haemonchus contortus and other ruminant trichostrongyloids. *Int. J. Parasitol.* 33 (11), 1129–1137. doi: 10.1016/S0020-7519(03)00167-X

Laing, R., Kikuchi, T., Martinelli, A., Tsai, I. J., Beech, R. N., Redman, E., et al. (2013). The genome and transcriptome of Haemonchus contortus, a key model parasite for drug and vaccine discovery. *Genome Biol.* 14 (8), R88. doi: 10.1186/gb-2013-14-8-r88

Laing, R., Martinelli, A., Tracey, A., Holroyd, N., Gilleard, J. S., and Cotton, J. A. (2016). Haemonchus contortus: genome structure, organization and comparative genomics. *Adv. Parasitol.* 93, 569–598. doi: 10.1016/bs.apar.2016.02.016

Leathwick, D. M. (2013). Managing anthelmintic resistance—parasite fitness, drug use strategy and the potential for reversion towards susceptibility. *Vet Parasitol.* 198 (1–2), 145–153. doi: 10.1016/j.vetpar.2013.08.022

Loukas, A., Bethony, J. M., Williamson, A. L., Goud, G. N., Mendez, S., Zhan, B., et al. (2004). Vaccination of dogs with a recombinant cysteine protease from the intestine of canine hookworms diminishes the fecundity and growth of worms. *J. Infect. Dis.* 189 (10), 1952–1961. doi: 10.1086/386346

Martin, J., Rosa, B. A., Ozersky, P., Hallsworth-Pepin, K., Zhang, X., Bhonagiri-Palsikar, V., et al. (2015). Helminth.net: expansions to Nematode.net and an introduction to Trematode.net. *Nucleic Acids Res.* 43, D698–706. doi: 10.1093/nar/gku1128

McCoy, C. J., Warnock, N. D., Atkinson, L. E., Atcheson, E., Martin, R. J., Robertson, A. P., et al. (2015). RNA interference in adult Ascaris suum—an opportunity for the development of a functional genomics platform that supports organism-, tissue- and cell-based biology in a nematode parasite. *Int. J. Parasitol. Parasites. Wildl.* 45 (11), 673–678. doi: 10.1016/j.ijpara.2015.05.003

McGhee, J. D., Fukushige, T., Krause, M. W., Minnema, S. E., Goszczynski, B., Gaudet, J., et al. (2009). ELT-2 is the predominant transcription factor controlling differentiation and function of the C. elegans intestine, from embryo to adult. *Dev. Biol.* 327 (2), 551–565. doi: 10.1016/j.ydbio.2008.11.034

Mitchell, A. L., Attwood, T. K., Babbitt, P. C., Blum, M., Bork, P., Bridge, A., et al. (2019). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* 47, D351–D360. doi: 10.1093/nar/gky1100

Morris, C. P., Bennuru, S., Kropp, L. E., Zweben, J. A., Meng, Z., Taylor, R. T., et al. (2015). A proteomic analysis of the body wall, digestive tract, and reproductive tract of Brugia malayi. *PLoS Negl. Trop. Dis.* 9 (9), e0004054. doi: 10.1371/journal.pntd.0004054

Munn, E. A., and Greenwood, C. A. (1984). The occurrence of submicrovillar endotube (modified termina web) and associated cytoskeletal structures in the intestinal epithelia of nematodes. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 306 (1125), 1–18. doi: 10.1098/rstb.1984.0079

O'Neill, M., Geary, J. F., Agnew, D. W., Mackenzie, C. D., and Geary, T. G. (2015). *In vitro* flubendazole-induced damage to vital tissues in adult females of the filarial nematode Brugia malayi. *Int. J. Parasitol. Drugs Drug Resist.* 5 (3), 135–140. doi: 10.1016/j.ijpddr.2015.06.002

O'Neill, M., Mansour, A., DiCosty, U., Geary, J., Dzimianski, M., McCall, S. D., et al. (2016). An *in vitro/in vivo* model to analyze the effects of flubendazole exposure on adult female Brugia malayi. *PLoS Negl. Trop. Dis.* 10 (5), e0004698. doi: 10.1371/journal.pntd.0004698

Perry, B., and Grace, D. (2009). The impacts of livestock diseases and their control on growth and development processes that are pro-poor. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 364 (1530), 2643–2655. doi: 10.1098/rstb.2009.0097

Piano, F., Schetter, A. J., Mangone, M., Stein, L., and Kemphues, K. J. (2000). RNAi analysis of genes expressed in the ovary of Caenorhabditis elegans. *Curr. Biol.* 10 (24), 1619–1622. doi: 10.1016/S0960-9822(00)00869-1

Piano, F., Schetter, A. J., Morton, D. G., Gunsalus, K. C., Reinke, V., Kim, S. K., et al. (2002). Gene clustering based on RNAi phenotypes of ovary-enriched genes in C. elegans.. *Curr. Biol.* 12 (22), 1959–1964. doi: 10.1016/S0960-9822(02)01301-5

Pillai, A., Ueno, S., Zhang, H., and Kato, Y. (2003). Induction of ASABF (Ascaris suum antibacterial factor)-type antimicrobial peptides by bacterial injection:

novel members of ASABF in the nematode Ascaris suum. *Biochem. J.* 371 (Pt 3), 663–668. doi: 10.1042/bj20021948

Pullan, R. L., Smith, J. L., Jasrasaria, R., and Brooker, S. J. (2014). Global numbers of infection and disease burden of soil transmitted helminth infections in 2010. *Parasit. Vectors* 7, 37. doi: 10.1186/1756-3305-7-37

Rawlings, N. D., and Morton, F. R. (2008). The MEROPS batch BLAST: a tool to detect peptidases and their non-peptidase homologues in a genome. *Biochimie.* 90 (2), 243–259. doi: 10.1016/j.biochi.2007.09.014

Rosa, B. A., Jasmer, D. P., and Mitreva, M. (2014). Genome-wide tissue-specific gene expression, co-expression and regulation of co-expressed genes in adult nematode Ascaris suum. *PLoS Negl. Trop. Dis.* 8 (2), e2678. doi: 10.1371/journal.pntd.0002678

Rosa, B. A., McNulty, S. N., Mitreva, M., and Jasmer, D. P. (2017). Direct experimental manipulation of intestinal cells in Ascaris suum, with minor influences on the global transcriptome. *Int. J. Parasitol.* 47 (5), 271–279. doi: 10.1016/j.ijpara.2016.12.005

Rosa, B. A., Townsend, R., Jasmer, D. P., and Mitreva, M. (2015). Functional and phylogenetic characterization of proteins detected in various nematode intestinal compartments. *Mol. Cell Proteomics* 14 (4), 812–827. doi: 10.1074/mcp.M114.046227

Selkirk, M. E., Huang, S. C., Knox, D. P., and Britton, C. (2012). The development of RNA interference (RNAi) in gastrointestinal nematodes. *Parasitology* 139 (5), 605–612. doi: 10.1017/S0031182011002332

Shao, H., Li, X., and Lok, J. B. (2017). Heritable genetic transformation of Strongyloides stercoralis by microinjection of plasmid DNA constructs into the male germline. *Int. J. Parasitol.* 47 (9), 511–515. doi: 10.1016/j.ijpara.2017.04.003

Shompole, S., Yao, C., Cheng, X., Knox, D., Johnson, S., and Jasmer, D. P. (2002). Distinct characteristics of two intestinal protein compartments discriminated by using fenbendazole and a benzimidazole resistant isolate of Haemonchus contortus.. *Exp. Parasitol.* 101 (4), 200–209. doi: 10.1016/S0014-4894(02)00135-2

Simonis, N., Rual, J. F., Carvunis, A. R., Tasan, M., Lemmens, I., Hirozane-Kishikawa, T., et al. (2009). Empirically controlled mapping of the Caenorhabditis elegans protein-protein interactome network. *Nat. Methods* 6 (1), 47–54. doi: 10.1038/nmeth.1279

Smith, T. S., Munn, E. A., Graham, M., Tavernor, A. S., and Greenwood, C. A. (1993). Purification and evaluation of the integral membrane protein H11 as a protective antigen against Haemonchus contortus. *Int. J. Parasitol.* 23 (2), 271–280. doi: 10.1016/0020-7519(93)90150-W

Smith, W. D., Newlands, G. F., Smith, S. K., Pettit, D., and Skuce, P. J. (2003a). Metalloendopeptidases from the intestinal brush border of Haemonchus contortus as protective antigens for sheep. *Parasite Immunol.* 25 (6), 313–323. doi: 10.1046/j.1365-3024.2003.00637.x

Smith, W. D., Skuce, P. J., Newlands, G. F., Smith, S. K., and Pettit, D. (2003b). Aspartyl proteases from the intestinal brush border of Haemonchus contortus as protective antigens for sheep. *Parasite Immunol.* 25 (11-12), 521–530. doi: 10.1111/j.0141-9838.2004.00667.x

Smith, W. D., Smith, S. K., and Murray, J. M. (1994). Protection studies with integral membrane fractions of Haemonchus contortus. *Parasite Immunol.* 16 (5), 231–241. doi: 10.1111/j.1365-3024.1994.tb00345.x

Sönnichsen, B., Koski, L. B., Walsh, A., Marschall, P., Neumann, B., Brehm, M., et al. (2005). Full-genome RNAi profiling of early embryogenesis in Caenorhabditis elegans. *Nature* 434 (7032), 462–469. doi: 10.1038/nature03353

The Gene Ontology Consortium. (2019). The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* 47 (D1), D330–D338. doi: 10.1093/nar/gky1055

Wang, J., Czech, B., Crunk, A., Wallace, A., Mitreva, M., Hannon, G. J., et al. (2011). Deep small RNA sequencing from the nematode Ascaris reveals conservation, functional diversification, and novel developmental profiles. *Genome Res.* 21 (9), 1462–1477. doi: 10.1101/gr.121426.111

Wang, J., Gao, S., Mostovoy, Y., Kang, Y., Zagoskin, M., Sun, Y., et al. (2017). Comparative genome analysis of programmed DNA elimination in nematodes. *Genome Res.* 27 (12), 2001–2014. doi: 10.1101/gr.225730.117

Wang, Q., Rosa, B. A., Jasmer, D. P., and Mitreva, M. (2015). Pan-Nematoda transcriptomic elucidation of essential intestinal functions and therapeutic targets with broad potential. *EBioMedicine* 2 (9), 1079–1089. doi: 10.1016/j.ebiom.2015.07.030

Wang, T., Van Steendam, K., Dhaenens, M., Vlaminck, J., Deforce, D., Jex, A. R., et al. (2013a). Proteomic analysis of the excretory-secretory products from larval stages of Ascaris suum reveals high abundance of glycosyl hydrolases. *PLoS Negl. Trop. Dis.* 7 (10), e2467. doi: 10.1371/journal.pntd.0002467

Wang, Z., Gao, X., Martin, J., Yin, Y., Abubucker, S., Rash, A. C., et al. (2013b). Gene expression analysis distinguishes tissue-specific and gender-related functions among adult Ascaris suum tissues. *Mol. Genet. Genomics* 288 (5–6), 243–260. doi: 10.1007/s00438-013-0743-y

Williams, A. R., Ramsay, A., Hansen, T. V., Ropiak, H. M., Mejer, H., Nejsum, P., et al. (2015). Anthelmintic activity of trans-cinnamaldehyde and A- and B-type proanthocyanidins derived from cinnamon (Cinnamomum verum). *Sci. Rep.* 5, 14791. doi: 10.1038/srep14791

Williamson, A. L., Brindley, P. J., Abbenante, G., Datu, B. J., Prociv, P., Berry, C., et al. (2003a). Hookworm aspartic protease, Na-APR-2, cleaves human hemoglobin and serum proteins in a host-specific fashion. *J. Infect. Dis.* 187 (3), 484–494. doi: 10.1086/367708

Williamson, A. L., Brindley, P. J., Knox, D. P., Hotez, P. J., and Loukas, A. (2003b). Digestive proteases of blood-feeding nematodes. *Trends Parasit.* 19 (9), 417–423. doi: 10.1016/S1471-4922(03)00189-2

Williamson, A. L., Lecchi, P., Turk, B. E., Choe, Y., Hotez, P. J., McKerrow, J. H., et al. (2004). A multi-enzyme cascade of hemoglobin proteolysis in the intestine of blood-feeding hookworms. *J. Biol. Chem.* 279 (34), 35950–35957. doi: 10.1074/jbc.M405842200

Yin, Y., Martin, J., Abubucker, S., Scott, A. L., McCarter, J. P., Wilson, R. K., et al. (2008). Intestinal transcriptomes of nematodes: comparison of the parasites Ascaris suum and Haemonchus contortus with the free-living Caenorhabditis elegans *PLoS Negl. Trop. Dis.* 2 (8), e269. doi: 10.1371/journal.pntd.0000269

Zintz, K., and Frank, W. (1982). Ultrastructural modifications in Heterakis spumosa after treatment with febantel or mebendazole. *Vet. Parasitol.* 10 (1), 47–56. doi: 10.1016/0304-4017(82)90006-1

Revisiting the Phylogenetic History of Helminths Through Genomics, the Case of the New *Echinococcus oligarthrus* Genome

Lucas L. Maldonado[1]*, Juan Pablo Arrabal[2], Mara Cecilia Rosenzvit[1],
Guilherme Corrêa De Oliveira[3] and Laura Kamenetzky[1]*

[1] IMPaM, CONICET, Facultad de Medicina, Universidad de Buenos Aires, Buenos Aires, Argentina,
[2] INMet, Instituto Nacional de Medicina Tropical, Puerto Iguazú, Argentina, [3] Instituto Tecnológico Vale, Belém, Brazil

The first parasitic helminth genome sequence was published in 2007; since then, only ~200 genomes have become available, most of them being draft assemblies. Nevertheless, despite the medical and economical global impact of helminthic infections, parasite genomes in public databases are underrepresented. Recently, through an integrative approach involving morphological, genetic, and ecological aspects, we have demonstrated that the complete life cycle of *Echinococcus oligarthrus* (Cestoda: Taeniidae) is present in South America. The neotropical *E. oligarthrus* parasite is capable of developing in any felid species and producing human infections. Neotropical echinococcosis is poorly understood yet and requires a complex medical examination to provide the appropriate intervention. Only a few cases of echinococcosis have been unequivocally identified and reported as a consequence of *E. oligarthrus* infections. Regarding phylogenetics, the analyses of mitogenomes and nuclear datasets have resulted in discordant topologies, and there is no unequivocal taxonomic classification of *Echinococcus* species so far. In this work, we sequenced and assembled the genome of *E. oligarthrus* that was isolated from agoutis (*Dasyprocta azarae*) naturally infected and performed the first comparative genomic study of a neotropical *Echinococcus* species. The *E. oligarthrus* genome assembly consisted of 86.22 Mb which showed ~90% identity and 76.3% coverage with *Echinococcus multilocularis* and contained the 85.0% of the total expected genes. Genetic variants analysis of whole genome revealed a higher rate of intraspecific genetic variability (23,301 SNPs; 0.22 SNPs/kb) rather than for the genomes of *E. multilocularis* and *Echinococcus canadensis* G7 but lower with respect to *Echinococcus granulosus* G1. Comparative genomics against *E. multilocularis*, *E. granulosus* G1, and *E. canadensis* G7 revealed 38,762, 125,147, and 170,049 homozygous polymorphic sites, respectively, indicating a higher genetic distance between *E. oligarthrus* and *E. granulosus sensu lato* species. The SNP distribution in chromosomes revealed a higher SNP density in the longest chromosomes. Phylogenetic analysis using whole-genome SNPs demonstrated that *E. oligarthrus* is one of the basal species of the genus *Echinococcus* and is phylogenetically closer to *E. multilocularis*. This

work sheds light on the *Echinococcus* phylogeny and settles the basis to study sylvatic *Echinococcus* species and their developmental evolutionary features.

HIGHLIGHTS

- Whole-genome SNP analysis showed high divergence among *Echinococcus sensu lato*.
- *Echinococcus* chromosomes have differential distribution of single-nucleotide polymorphisms.
- Sylvatic *Echinococcus* species are phylogenetically closer to each other than *Echinococcus granulosus sensu lato* species.
- Phylogenetic analysis demonstrated that *Echinococcus oligarthrus* is one of the basal species.

INTRODUCTION

Helminth parasites are a highly diverse group that involves many parasites of biomedical, veterinary, and economic importance including roundworms (nematodes) and flatworms (Platyhelminthes: trematodes and cestodes). Taxonomic assignment of the parasitic helminth taxa is a particularly arduous task. Indeed, parasites are typically difficult to culture and analyze independently of their hosts and the parasite body fossil samples are scarce due to their small size, lack of hard parts, and their lifestyle within the host. In addition, molecular analyses do not often include all the species or include only partial sequences (De Baets et al., 2015). Among Platyhelminthes, the classification and nomenclature within the genus *Echinococcus* is a controversial topic (McManus, 2013; Lymbery et al., 2015). The strobilar stage of these parasites occurs in the small intestine of a definitive carnivore host; the metacestodes develop in the organs of an herbivorous intermediate host that is the prey of the final hosts. In the last decades, several subspecies of *Echinococcus granulosus* were proposed based mainly on the intermediate host specificity (Verster, 1969). However, most of the early proposed subspecific taxa were relegated to synonymy under the name *E. granulosus* due to their sympatric distributions and because they are indistinguishable at the morphological level (Rausch et al., 1967). Consequently, based on the production of a distinctive form of echinococcosis in humans, only four species had been retained: *E. granulosus sensu lato*, which causes cystic echinococcosis; *Echinococcus oligarthrus*, which causes unicystic echinococcosis; *Echinococcus vogeli*, which causes polycystic echinococcosis; and *Echinococcus multilocularis*, which causes alveolar echinococcosis. The pathogenicity degree of the echinococcosis infections depends on the characteristics of the metacestode development, which is different in each one of the four species mentioned before (Lymbery, 2017). Nowadays, genetic and genomic mitochondrial analysis allows the revision of the phylogeny of the *Echinococcus* genus determining the species rank for nine taxa: *E. granulosus*

sensu stricto, *Echinococcus canadensis*, *Echinococcus ortleppi*, *Echinococcus equinus*, *Echinococcus felidis*, *E. oligarthrus*, *E. vogeli*, *E. multilocularis*, and *Echinococcus shiquicus* (Xiao et al., 2006; Nakao et al., 2007; Hüttner et al., 2008; Nakao et al., 2013a; Kinkar et al., 2017). Nuclear DNA has also been used to reconstruct the phylogeny of taeniid parasites, which differs from the phylogeny obtained with mitochondrial data (Saarma et al., 2009; Knapp et al., 2011). Even though there are certain common features in the taxonomy regardless of the origin of the molecular data. *E. felidis* and *E. granulosus s. s.* are sister species, and *E. ortleppi* is closely related to the different genotypes of *E. canadensis*. However, the position of the neotropical species, *E. vogeli* and *E. oligarthrus* (basal or non-basal), and whether *E. multilocularis* and *E. shiquicus* are sister species remain unknown (Lymbery, 2017). Several authors agree that further analyses using more nuclear DNA sequences are required in order to completely resolve the relationships among putative species within the genus (Saarma et al., 2009; Nakao et al., 2013a; Lymbery, 2017). Particularly, neotropical species have been the least studied and only a few cases have been reported and published so far (D'Alessandro and Rausch, 2008; Soares et al., 2013; Arrabal et al., 2017). Hence, a better sampling and understanding of the neotropical species will help to resolve the *Echinococcus* phylogeny. In 2013, with the publication of the first tapeworm genomes that included *E. multilocularis* and *E. granulosus* G1 species (Tsai et al., 2013; Zheng et al., 2013), the "Tapeworm genome era" began. In 2017, we sequenced and assembled the *E. canadensis* G7 genome and performed several comparative genomic analyses confirming the species status of the taxa (Maldonado et al., 2017). In this work, with the aim of obtaining a deeper description of the *Echinococcus* phylogeny, we sequenced the complete *E. oligarthrus* genome and performed whole-genome variant analysis among all the *Echinococcus* species that are currently available. The results presented here demonstrate the basal origin of the neotropical *Echinococcus* species and propose that the use of complete genome data is crucial for the unequivocal helminth phylogenetic studies.

MATERIALS AND METHODS

Sample Collection, DNA Extraction, and Next-Generation Sequencing
Parasites Material

E. oligarthrus cysts were collected from Iguazú National Park, in the North of Misiones province, Argentina. Cysts were obtained from the livers of naturally infected agoutis (*Dasyprocta azarae*). The animals involved in this study were not subjected to any experimental procedure. All the samples used in this study were

collected post-mortem from road-killed animals. For genome sequencing purposes, protoscoleces were aseptically removed from the cysts and extensively washed in phosphate buffer saline and visualized under an optical microscope. The species and genotype were determined by sequencing a fragment of the mitochondrial cytochrome C oxidase subunit 1 (COX1) (Arrabal et al., 2017).

DNA Isolation, Library Construction, and DNA Sequencing

The isolation of high-quality genomic DNA was performed by the phenol/chloroform method as previously described (Maldonado et al., 2017). Briefly, the samples were quantified using a Qubit Fluorometer (Invitrogen), the quality was evaluated by the OD rate 260/280 and 260/230 using a NanoDrop (ThermoFisher Scientific). MiSeq Illumina libraries were prepared as follows. For each library preparation, 50 ng of DNA was subjected to a random tagmentation reaction, and DNA was simultaneously fragmented and linked to specific adapters using the Nextera® XT DNA Sample Preparation Kit, according to the manufacturer's instructions. Two libraries of 530-bp fragment size were obtained and subjected to 500 sequencing cycles (2 × 250 bp) using the MiSeq v2 Reagent Kit. The quality of the Illumina reads was evaluated with FastQC v0.10.1, and the reads were trimmed and end-clipped to a Phred score of 33 using Trimmomatic (Bolger et al., 2014).

De Novo Assembly of *E. oligarthrus* NGS Reads

The genome of *E. oligarthrus* was assembled from a combination of two paired-end libraries sequenced in the Illumina MiSeq platform. The genome assembly involved several steps. First, a preliminary *de novo* assembly using SPAdes 3.6 (Bankevich et al., 2012; Safonova et al., 2014) was performed. The assembly sequences were screened against the NCBI nt database, using Nucleotide–Nucleotide BLAST v2.6.0+ (available at ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/nt.gz) in megablast mode, with an e-value cutoff of 1e−25 and a culling limit of 2 and using DIAMOND tblastx against SwissProt (Buchfink et al., 2015). Raw, paired-end Illumina reads were mapped against the assembly using Bowtie2 (Langmead and Salzberg, 2012). The output was converted to a BAM file using Samtools (Li and Durbin, 2009). Blobtools v1.1 (Laetsch and Blaxter, 2017) was used to create taxon-annotated GC-coverage plots for *E. oligarthrus* genome assembly and to identify the target sequences and target reads using as input the Nucleotide–Nucleotide and DIAMOND BLAST (Buchfink et al., 2015) and the raw read mapping results. Sequences that did not match the Platyhelminthes taxon as a top BLAST hit at the phylum level were filtered out. After removing contaminants, sequences that did match as a top BLAST hit at the phylum level and whose base coverage was >4× were used to recover the target reads. Before the assembly, we used Jellyfish to create a k-mer histogram and the convergence was analyzed with genomescope in order to evaluate whether after cleaning with blobtools the remaining reads could be used to obtain an acceptable assembly. The target reads extracted from the bam files were used to re-assemble the *E. oligarthrus* genome using SPAdes 3.12 (Bankevich et al., 2012; Safonova et al., 2014). The scaffolds of *E. oligarthus* were obtained with

Chromosomer (Tamazian et al., 2016) and using *E. multilocularis* as the reference genome [WormBase ParaSite, Version WBPS12 (WS267)]. Redundant unplaced and unlocalized sequences were discarded from the final assembly. The fragment sequence was considered unplaced if two or more alignments located on different reference sequences or unlocalized if two or more alignments located on the same reference sequences according to Tamazian et al. (2016). The contigs that did not map to the reference genome were included in the final assembly. The standard quality metrics of the assembly such as N50, the total number of contigs and the total length of the assembly were evaluated using QUAST (Gurevich et al., 2013). Putative non-target contigs shorter than 500 bp in length were removed from the final assembly. The completeness of the gene space was validated using BUSCO2 (Simão et al., 2015) and eukaryote CEGGs database. Furthermore, the core of cestodes genes [genes contained in all cestode species according to Maldonado et al. (2017)] was screened using BLAST. Coverage and depth coverage were calculated with custom scripts. Depth coverage refers to the number of times that the same region or position in the reference genome is represented by the assembled genome. Coverage refers to the percentage of the total length of the reference genome that is represented by the assembled genome.

Gene Prediction and Annotation

The gene annotation of *E. oligarthrus* was performed by transferring the gene annotations from the *E. multilocularis* genome [WormBase ParaSite, Version WBPS12 (WS267)] using a manual and own scripting approach. CDS and protein sequences of *E. multilocularis* and the scaffolds of *E. oligarthrus* were used for this purpose. First, the gene allocated regions were identified using BLAST with an e-value cutoff of 1e−10 and one best target hit. The CDS fragments of *E. oligarthrus* were extracted using bedtools (Quinlan and Hall, 2010; Quinlan, 2014) and the transcripts were re-assembled using Chromosomer (Tamazian et al., 2016). GeneWise (Birney et al., 2004) was used to find the correct frameshift of the CDS and to obtain the final datasets of CDS and proteins. The coordinates of the annotations were assessed using Exonerate (Slater et al., 2005) and added to a final GFF file. The performance of gene annotation and basic statistics for *E. oligarthrus* gene models, including the average intron/exon lengths and the number of introns, were calculated using Eval (Keibler and Brent, 2003). The functional gene annotation was performed using InterProScan-5.7-48.0 (Quevillon et al., 2005) and InterPro2GO databases were used to assign Gene Ontology (GO) terms (Ashburner et al., 2000). Gene models were subjected to BLAST search (Stemmer et al., 2013) against UniprotDB and Blast2GO (https://www.blast2go.com/) was used to define the final annotations and GO terms stats. The GO terms were analyzed using GO.db database implementing R, Bioconductor version: Release (3.8) with custom scripts. Proteins studied in this work were searched using BLAST (Stemmer et al., 2013) against UniProtKB/Swiss-Prot databases. Protein domains were screened against PFAM, and Prosite databases using PFAM_scan (Finn, 2006) or HMMscan 3.0. Global pairwise alignments were performed using Needle software (Rice et al., 2000) in order to identify the orthologous genes related to host–parasite

interactions between *E. oligarthrus* and *E. multilocularis* (Brehm and Koziol, 2017). Identity and coverage stats were calculated and only hits with more than 50% of coverage were selected.

Phylogeny With Nuclear Molecular Markers

Molecular markers of the *Echinococcus* nuclear genome were downloaded from GenBank (Kinkar et al., 2017). For each locus, homologous regions were extracted from complete *Echinococcus* genomes [WormBase ParaSite, Version WBPS12 (WS267)] and the *E. oligarthrus* genome obtained in this work using custom scripts. Homologous genomic sequences from *Taenia solium* were employed as outgroup. The DNA sequences were aligned with ClustalX (v2.0.12) and multiple alignments were edited with BioEdit (v7.1.3). Phylogenetic analyses of concatenated nuclear markers were performed using the maximum likelihood method and Tamura-Nei model implemented by MEGAX software (Kumar et al., 2018). The bootstrap consensus trees inferred from 1000 replicates were retained. The percentage of replicate trees in which the associated taxa clustered together in more than 50% in the bootstrap test are shown for each node. The concatenated data resulted in 1552 bp length for seven taxa. A second phylogeny was calculated using MrBayes 3.1.2 (Ronquist et al., 2012). The evolutionary model was set on generalized time reversible substitution model with gamma-distributed rate variation across the sites and a proportion of invariable sites (GTR +G + I model) with at least 200 samples from the posterior probability distribution, and diagnostics calculated every 1000 generations. We use an exponential prior on the branch length with mean = 0.1 substitutions/site.

Phylogeny Using Whole-Genome SNPs and Artificial Genome Sequence Construction

Variant calling was performed as described in the "Whole-Genome SNP analysis" section. Genome-wide SNPs were used to perform phylogeny analysis as follows: first, heterozygous SNPs were removed and only the homozygous SNPs with a depth coverage >20× and strictly covered in all the *Echinococcus* species were selected to correct for complete lineage sorting. Afterwards, the homozygous SNP loci were concatenated and the resulting sequence alignment was used to create a phylogenetic tree by implementing the Maximum Likelihood and Bayesian method as above. PartitionFinder 2 (Lanfear et al., 2016) was used to select the best-fit partitioning schemes and models of evolution for the phylogenetic analysis. The selected model was implemented to perform phylogeny analysis.

Whole-Genome SNP Analysis

Genome sequences of *Echinococcus* species were ordered in chromosomes according to the last version of the *E. multilocularis* genome [WormBase ParaSite, Version WBPS12 (WS267)] using Chromosomer (Tamazian et al., 2016) and were used as mapping templates for the variant calling analyses. Variant calling was performed as follows: first, all of the raw reads from *E. canadensis* G7, *E. granulosus* G1, *E. multilocularis*, and *E. oligarthrus* libraries

were processed and filtered by quality. Then, the reads were first mapped against their own reference genomes and then against the genomes of the other three *Echinococcus* species using bowtie2 (Langmead and Salzberg, 2012). Mapping statistics were calculated with Bamtools (Barnett et al., 2011), and duplications were marked and discarded using picard-tools v-2.18 (http://broadinstitute. github.io/picard/). The variant calling was performed using bcftools (Narasimhan et al., 2016; Danecek and McCarthy, 2017) and GATK (McKenna et al., 2010) using the following parameters: variation frequency was set >40% with a depth coverage of at least 70% of the total mean coverage; the base quality of both reference site and variation site was set to >30. Insertion and deletions (indels) were filtered out with VCFtools (Danecek et al., 2011), and SNPs with less than 10 bp far from indels were removed to avoid false-positive SNP. In order to annotate the heterozygous and homozygous polymorphic sites, the reads were first mapped against their own reference genomes and then against the other target genomes. Heterozygous sites were retained only if both forward and reverse reads mapped against the reference and alternative allele at a given nucleotide position whose depth coverage was at least 70% of the total mean coverage supporting that position. Homozygous polymorphic sites were annotated if the forward and reverse reads mapped onto the alternative allele with at least 70% of the total mean coverage supporting that position and if there were no reads supporting the reference allele. Homozygous and heterozygous variant sites were registered for all of the species. The Transition/transversion ratios were calculated using VCFtools (Danecek et al., 2011), and the annotation and classification of SNPs based on the effect of annotated genes were carried out with SnpEff v4.0 (Cingolani et al., 2012). Graphics were built using R software. (https://www.r-project.org/).

RESULTS

The *E. oligarthrus* Genome

The *E. oligarthrus* genome was assembled from two paired-end libraries sequenced in Illumina MiSeq. High-quality genomic DNA was purified from two cysts isolated from its natural host, *D. azarae*, which was naturally infected and found in Iguazú National Park, Misiones province, Argentina. Microscopic and macroscopic findings indicated that it was a neotropical *Echinococcus* species. The species and the genotype were confirmed by polymerase chain reaction (PCR) amplification of cytochrome oxidase 1 (cox1) followed by direct sequencing. The *E. oligarthrus* group 1 cox1 sequence was determined (Arrabal et al., 2017). The *E. oligarthrus* genome assembly was performed using a *de novo* assembly strategy, and the best assembly was chosen based on its quality metrics. Due to the type of the metacestode development that consists of close interaction with the host's tissue and in spite of having done a careful extraction of the parasite material, we sequenced a high proportion of host's DNA. Indeed, preliminary analysis of raw reads revealed that the sequencing yield of parasite DNA was ~22%, whereas the remaining ~78% of the sequences was from presumably *D. azarae*, whose genome has not been sequenced yet. Therefore, we performed a screening of

potential contaminants before obtaining the final assembly. Since the genome of *D. azarae* is unknown and non-related sequences are available, we made a thorough identification of sequences truly derived from the target genome using Blob tools software (Laetsch and Blaxter, 2017). This software was specially designed for DNA contamination assessment and is particularly useful for organisms of parasitic origin. Here, we performed a taxonomic selection of the target sequences using the Platyhelminthes taxon (Taxonomy ID: 6157) and created a taxon-annotated GC-coverage plot that allowed us to identify Platyhelminthes contigs and discard all the non-target sequences (**Supplementary Figure 1**). After removing the contaminants, sequences that did match as a top BLAST hit at the phylum level and whose base coverage was >4× and %GC ~ 41% (typical GC content of *Echinococcus* genus) were used to recover the target reads and re-assemble the *E. oligarthrus* genome (see Materials and methods for more details). The final assembly contained 74513 contigs that were further used to obtain the genome scaffolds. Redundant unplaced and unlocalized sequences (19.7 Mb) were removed from the contigs to obtain the final scaffolds assembly (see Materials and methods for more details). The contigs that did not map to the reference genome (did not allocate in scaffolds) were included in the final assembly. The final *E. oligarthrus* genome assembly was composed of 3764 sequences comprising a total of 86.22 Mb with an average GC content of 41% and showed an N50 ~ 10 Mb and ~20× in depth coverage. The assembly metrics for contigs and scaffolds are shown in **Table 1** and **Supplementary Table S1**. The nuclear genome of *E. oligarthrus* showed 76.3% coverage and ~90% identity with the genome of *E. multilocularis* whose genome size is ~115 Mb [WormBase ParaSite, Version WBPS12 (WS267)]. The percentage average of nucleotide identity among *Echinococcus* chromosomes ranged from 88.7% to 92.7% (**Supplementary Table S2**). The assembly also included the *E. oligarthrus* mitochondrial genome composed of one scaffold that was obtained from the contigs Eoli_00665 and Eoli_02629. The scaffold length was 13,893 bp with 98% coverage and 96% nucleotide identity to the *E. oligarthrus* mitochondrial reference genome (GenBank accession number AB208545). As expected, the nucleotide identity of the mitochondrial genomes was lower in comparison with other *Echinococcus* species (**Supplementary Figure 2**). To assess the completeness of the genome assembly, we evaluated the gene space using "Benchmarking Universal Single-Copy Orthologs" (BUSCO) (Simão et al., 2015), which measures the genome completeness based on evolutionarily informed expectations of gene content. In this analysis, we identified 77.6% (235 of the 303 core genes) that are expected to be present in all metazoans, including 125 complete and duplicated, 110 fragmented, and 68 missing orthologs. The fragmented nature of the assembly may have prevented many genes from meeting the stringent matching criteria implemented by BUSCO. Indeed, the BLAST results suggest that most of the core genes are identifiable in the genome, even though many genes are present as fragments within the assembly. Also, due to the fragmentation level of the genome assembly, a hybrid and *ab initio* gene prediction is likely

TABLE 1 | Genome-wide statistics for the *E. oligarthrus* assembly and gene findings.

Genome statistics	Scaffolds	Contigs
Size of genome (Mb) (*)	86.2	105.9
GC content (%)	41	41
Number of sequences	3764	74,513
N50 (Mb)	10.2	12.83 kb
Gaps (Ns/100 kb)	1366	0
Largest contig (Mb)	16.0	63.7 kb
Deep coverage	20×	20×
Number of predicted genes	8753	
Gene density per Mb	101.5	
Length of proteome (amino acids)	2,009,586	
Maximum protein length (amino acids)	2254	
Average protein length (amino acids)	229	
Average exon length (bp)	203	
Median exon length (bp)	158	
Average exons per transcript	3	
Median exons per transcript	2	
Total length of contained introns (kb)	8622	
Average intron length (bp)	709	
Median intron length (bp)	262	
BUSCO	235/303	162/303

(*) *Redundant sequences (19.7 Mb) were removed from the contigs to obtain the final scaffolds assembly.*

impracticable. Since all the available gene annotation transfer tools tried here showed low efficiency, the gene annotation was performed by manual gene annotation transfer as described in the Materials and methods section. First, we used BLAST to search for the best suitable *Echinococcus* species to be used as a template. In this regard, we found that *E. multilocularis* was the most suitable species because of the integrity of the assembly and the higher identity with the genome of *E. oligarthrus* (**Supplementary Table S2**). The final set of *E. oligarthrus* genes comprised 8753 genes coding for proteins (4494 genes with coverage > 50% and 4259 genes with coverage < 50%) (**Supplementary Table S3**). The whole set of genes represented the 85.0% of the total genes expected to be found in species of the genus *Echinococcus*. The predicted proteins were also screened against the conserved core of genes that are present in all the cestodes species according to Maldonado et al. (2017) using BLAST. In this regard, 4872 out of 5203 genes were found for *E. oligarthrus* using e values of <1e−12, which comprised the 93.6% of the total conserved core of genes in cestodes, indicating that the genome contains useful molecular data (**Supplementary Table S4**).

E. oligarthrus Genes

GO terms were assigned to the 40% of *E. oligarthrus* proteins. In relation to the Molecular Function GO terms frequency, the two main categories found were "binding (GO:0005488)" and "catalytic activity (GO:0003824)", which is in accordance with the GO terms frequency observed in other cestode genomes (Hahn et al., 2014) and *Echinococcus* species (Maldonado et al., 2017). For the Biological Process GO term, the highest frequencies observed were the categories "cellular process (GO:0009987)" and "metabolic process (GO:0008152)", also in accordance with the GO terms frequency observed in the related organism (**Supplementary Figure 3**).

From the total gene-set, we searched for genes that have already been described as having a role in host–parasite interactions in *Echinococcus* (Brehm and Koziol, 2017). In this regard, we found 56 genes involved in host–parasite interactions whose coverages and identities to *E. multilocularis* orthologous genes were >50% (**Supplementary Table S5**). Particularly, two genes encoding for the epidermal growth factor (EGF) tyrosine kinase receptors (Eoli_000075800 and Eoli_000617300) with a high percentage of identity (59.8% and 52.6%) and coverage (65.3% and 51.4%) were identified. Also, we found a gene encoding for fibroblast growth factor (FGF) receptor tyrosine kinase (Eoli_000833200) whose percentage of identity and coverage values were 76% and 100%, respectively. Regarding nuclear receptor hormones, we found nine orthologous genes with high identity and coverage (83.8% and 51.8% on average, respectively) including the cestode-specific nuclear hormone receptor Eoli_000937000. Non-kinase receptors were also found, comprising four genes, including Frizzled G protein-coupled receptors (GPCR) (Eoli_000682100) with high identity and coverage (88.6% and 87.6%, respectively). Finally, we identified the amino acid transporters (DAACS family), lipid binding proteins (FABPs), and antigens (AgB and Eg95), specific and conserved proteins in cestodes/*Echinococcus*.

Comparative Whole Genome Based on SNPs Analysis

The nucleotide variation in the genomes of different *Echinococcus* species was assessed through variant calling analysis. Genetic variants were identified among the genomes of *E. oligarthrus*, *E. multilocularis*, *E. granulosus* G1, and *E. canadensis* G7. Here, we focused on the study of single-nucleotide polymorphisms (SNPs). In order to perform this analysis, NGS raw reads were mapped against a reference genome composed of chromosomes and contigs. For all the analyses, the reads were first mapped against their own reference genomes and then against the genome of the corresponding analyzed species. Homozygous and heterozygous variant sites were identified and were marked in both the reference and the alternative allele (see Materials and Methods for more details). First, we evaluated the intraspecific variation of the *Echinococcus*

genomes. As described in Maldonado et al. (2017). *E. granulosus* G1 genome exhibited the highest number of intraspecific variant sites (74,796 SNPs, 0.65 SNPs/kb), followed by *E. oligarthrus* (23,301 SNP, 0.23 SNP/kb), *E. canadensis* G7 (10,791 SNPs, 0.095 SNPs/kb), and *E. multilocularis* (1,287 SNPs, 0.011 SNPs/kb). With regard to the genetic diversity observed among the *Echinococcus* species, we observed that the SNP distribution was similar to the distribution described in our previous research (Maldonado et al., 2017). *E. canadensis* G7 and *E. granulosus* G1 showed a higher number of SNPs (842,322, Ts/Tv = 2.97) between each other than between *E. canadensis* G7 and *E. multilocularis* (314,176, Ts/Tv = 2.97), and in comparison to *E. granulosus* G1 and *E. multilocularis* (272,138, Ts/Tv = 2.94). Furthermore, the pair *E. canadensis* G7 and *E. granulosus* G1 also showed almost 10 orders of magnitude more SNPs than between *E. oligarthrus* and *E. multilocularis* (38,911, Ts/Tv = 3.25). The number of SNPs between *E. oligarthrus* and *E. granulosus* (126,472 Ts/Tv = 2.99) was similar to the number of SNPs between *E. oligarthrus* and *E. canadensis* G7 (171,135, Ts/Tv = 3.00). We also identified homozygous and heterozygous variant sites for the four *Echinococcus* species, for both the reference and the alternative allele in each case. The number of homozygous SNPs was 313,992 for *E. canadensis* G7 and *E. multilocularis*, 266,180 for *E. granulosus* G1 and *E. multilocularis*, 38,762 for *E. oligarthrus* and *E. multilocularis*, 830,768 for *E. canadensis* G7 and *E. granulosus* G1, 125,147 for *E. oligarthrus* and *E. granulosus* G1, and 170,049 for *E. oligarthrus* and *E. canadensis* G7 (**Table 2**). Moreover, the SNP density was assessed in terms of the number of SNPs per 1-Mb length of the *Echinococcus* chromosomes. For all the *Echinococcus* genomes, the highest SNP density (SNPs/Mb of the chromosome) was found for chromosome 1. The chromosomes 2, 3, and 4 showed a slightly lower SNP density than chromosome 1 and chromosomes 5, 6, 7, 8, and 9 showed the lowest SNP density (**Figure 1**). However, the intraspecific SNP density distribution for *E. granulosus* G1 was higher in chromosomes 1, 5, and 9. We also evaluated the SNP distribution of *E. oligarthrus* in the coding and non-coding genomic regions (intergenic, exons, and introns) of *E. multilocularis*. In this regard, the 38,911 SNPs distributed with a higher rate in exons and introns rather than in the intergenic regions. However, those distributed in the coding regions exhibited a higher rate of synonymous changes (67.3%) rather than missense changes (32.6%) (**Supplementary Figure 4**).

TABLE 2 | Number and type of SNPs among *Echinococcus* species.

Species	Sample	Alt : Het	Alt : Hom	Total number of SNPs	Ts/Tv
E. multilocularis	JAVA05	1152	135	1287	1.98
E. canadensis G7	PH14	9937	854	10,791	3.25
E. granulosus G1	GH09/3	69,672	5124	74,796	2.32
E. oligarthrus	ADA3-ADA5	23,074	227	23,301	2.97
E. canadensis G7 vs *E. multilocularis*[a]	PH14–JAVA05	168	313,992	314,176	2.97
E. granulosus G1 vs *E. multilocularis*[b]	GH09/3–JAVA05	5940	266,180	272,138	2.94
E. oligarthrus vs *E. multilocularis*[c]	ADA–JAVA05	147	38,762	38,911	3.25
E. canadensis vs *E. granulosus* G1[d]	PH14–H95/5	1359	830,768	842,322	2.97
E. oligarthrus vs *E. granulosus* G1[e]	ADA–H95/5	667	125,147	126,472	2.99
E. oligarthrus vs *E. canadensis* G7[f]	ADA3–ADA5–PH14	1044	170,049	171,135	3.00

[a]*Ref/Alt : Hom/Hom: 6; Ref/Alt : Hom/Het:0; Ref/Alt : Het/Het:1; Ref/Alt : Het/Hom:9.* [b]*Ref/Alt : Hom/Hom: 7; Ref/Alt : Hom/Het:0; Ref/Alt : Het/Het:0; Ref/Alt : Het/Hom:11.* [c]*Ref/Alt : Hom/Hom: 0; Ref/Alt : Hom/Het:0; Ref/Alt : Het/Het:0; Ref/Alt : Het/Hom:2.* [d]*Ref/Alt : Hom/Hom: 944; Ref/Alt : Hom/Het:0; Ref/Alt : Het/Het:13; Ref/Alt : Het/Hom:9238.* [e]*Ref/Alt : Hom/Hom: 55; Ref/Alt : Hom/Het:0; Ref/Alt : Het/Het:0; Ref/Alt : Het/Hom:603.* [f]*Ref/Alt : Hom/Hom: 7; Ref/Alt : Hom/Het:0; Ref/Alt : Het/Het:7; Ref/Alt : Het/Hom:35.*
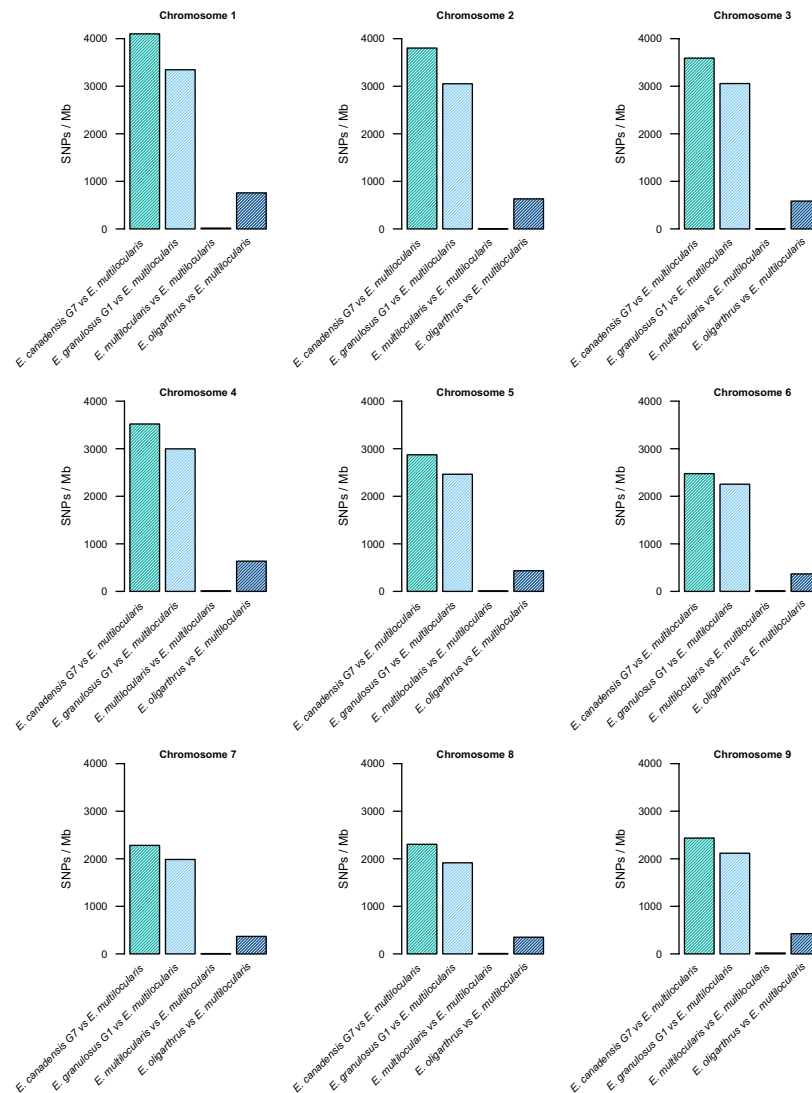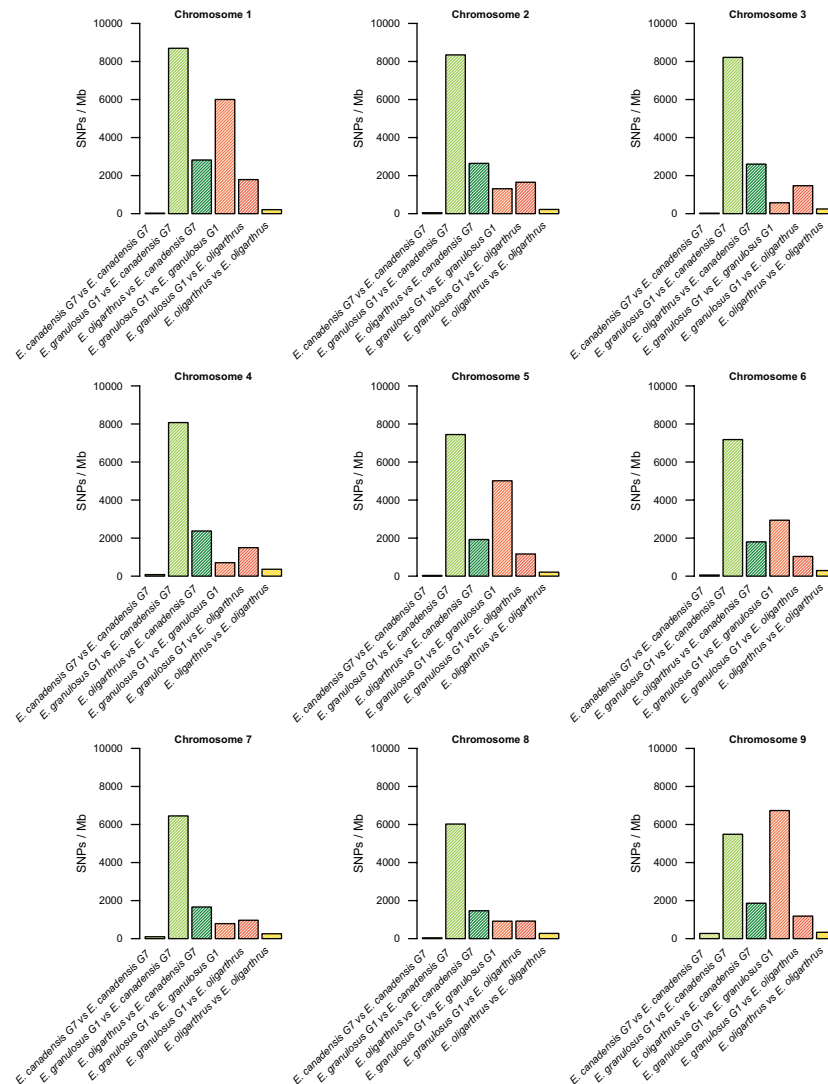
FIGURE 1 | Continued

Echinococcus Phylogeny Reconstruction by Whole-Genome SNP Analysis

In order to evaluate the contribution of the SNPs to the genetic diversity among the different *Echinococcus* species, we performed phylogenetic analyses by implementing three different approaches: the first used the whole genome variant sites for the four *Echinococcus* species (only the SNPs), which consisted of analyzing 244,246 sites in each genome arising a total of 2.08 Mb; the second approach used genome regions for the four *Echinococcus* species whose depth coverage was >20× and that strictly contained variant sites supported by more than 20 reads, which involved the analysis of 40,179,279 sites in each genome arising in a total of ~162 Mb. For the third approach, we used only coding regions for the four *Echinococcus* species whose depth coverage was >20× and that strictly contained variant sites supported by more than 20 reads. This involved the analysis

of 42,200 sites in each genome arising in a total of 168.8 kb. In all the cases, only the homozygous SNPs were used to perform phylogenetic analyses. The sites and sequences retained here were concatenated and the resulting alignment was used to create the phylogenetic trees by implementing the maximum likelihood and Bayesian methods (see Materials and methods for more details).

The construction of the phylogenetic tree implementing the Bayesian method and using whole-genome SNPs showed a topology that demonstrated a higher genetic distance between *E. canadensis* G7 and *E. granulosus* G1 in comparison with the common node from which *E. multilocularis* and *E. oligarthrus* diverge equidistantly at a very short distance (**Figure 2A**). *E. multilocularis* and *E. oligarthrus* exhibited a low genetic diversity between each other. PartitionFinder 2 (Lanfear et al., 2016) was used to select the best-fit partitioning schemes and models of evolution for the phylogenetic analysis. Transversional substitution
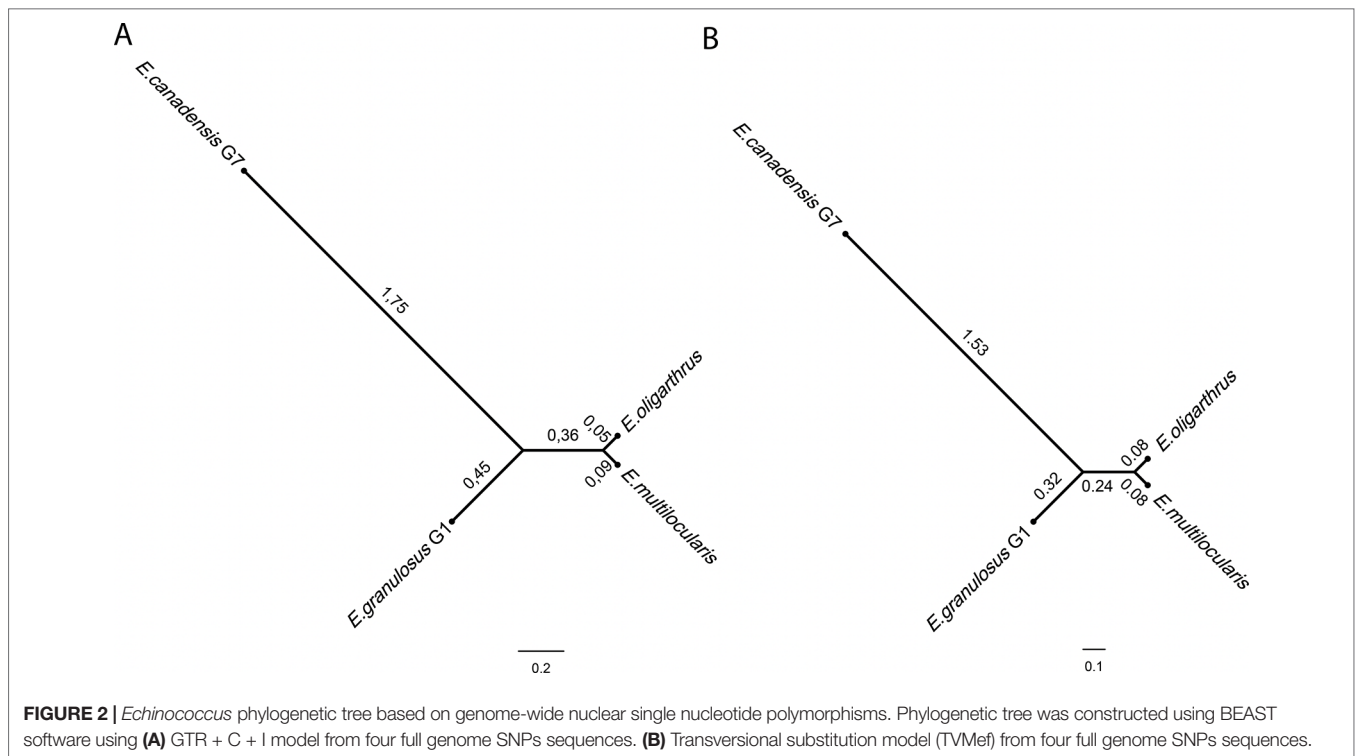
FIGURE 1 | Single-nucleotide polymorphisms in *Echinococcus* chromosomes. Density of single-nucleotide polymorphisms (SNPs/Mb) by chromosome 1–9 among *E. oligarthrus*, *E. multilocularis*, *E. canadensis* G7, and *E. granulosus* G1 species.

model (TVMef) was the best-fitted model. Phylogenetic analysis using TVMef was consistent with the previous result (**Figure 2B**). The topologies of the trees were also consistent with the topology observed using the Maximum Likelihood method (**Supplementary Figure 5A**). Phylogenetic analysis using genomic regions with depth coverage >20× sites and coding regions with depth coverage >20× also showed the same topology, except for the branches that exhibited different lengths (**Supplementary Figures 5B, C**). The total length of the coding sequences used here was ~42.2 kb and the genes sampled and used for this purpose are listed in **Supplementary Table S6**.

In addition, we registered how many polymorphic loci are shared among the *Echinococcus* species containing the same polymorphism. In previous research, we determined that the number of shared loci was higher when we used *E. multilocularis* as the reference genome, demonstrating to be the basal species
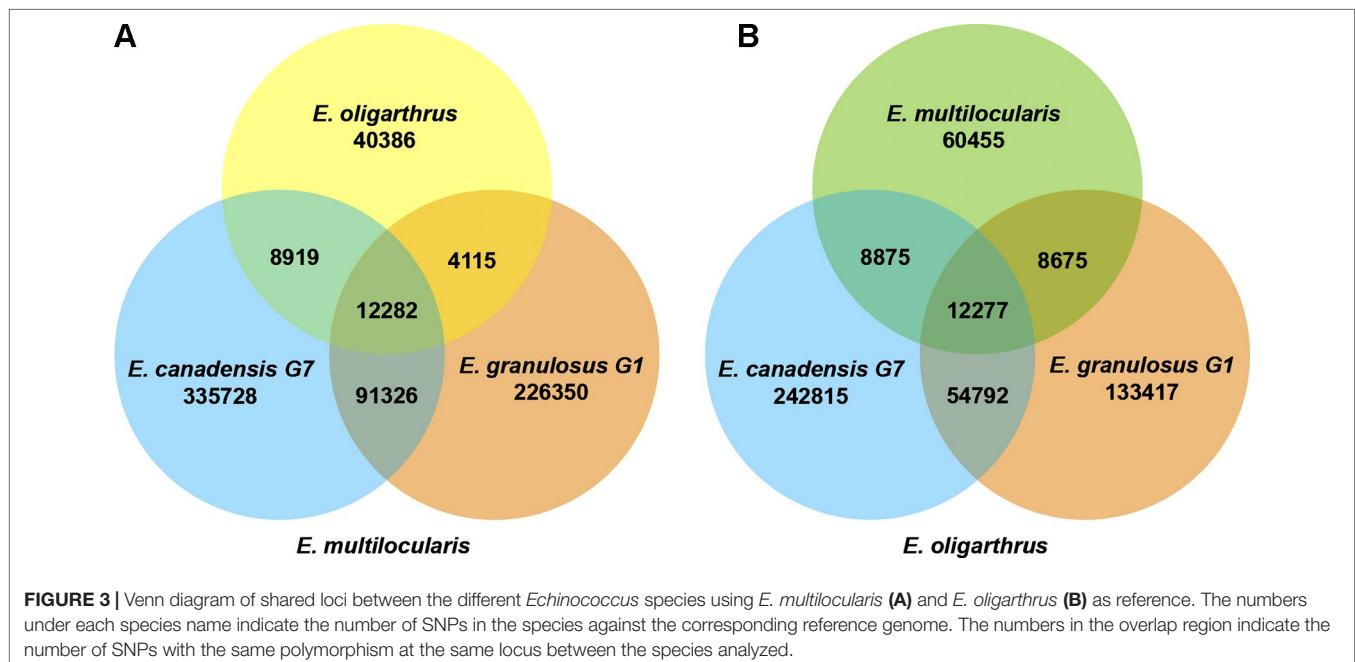
and discarding *E. granulosus* G1 and *E. canadensis* G7 as possible candidates (Maldonado et al., 2017). In order to evaluate whether *E. oligarthrus* occupies a basal position in the genus, we incorporated *E. oligarthrus* and analyzed the number of shared loci using both *E. multilocularis* and *E. oligarthrus* as reference genomes and compared the results to each other. Due to the different genome quality assemblies and coverage, we normalized the number of SNPs in shared loci using the effective length of the sampled regions under the assumption that all the genome regions are equally subjected to mutation. In this regard, we found that 12,282 loci shared the same nucleotide change among *E. canadensis* G7, *E. granulosus* G1, and *E. oligarthrus* with respect to *E. multilocularis* (*E. multilocularis* used as reference). Moreover, the number of shared loci between *E. oligarthrus* and *E. granulosus* G1, between *E. oligarthrus* and *E. canadensis* G7, and between *E. canadensis* G7 and *E. granulosus* G1 was 4115,

FIGURE 2 | *Echinococcus* phylogenetic tree based on genome-wide nuclear single nucleotide polymorphisms. Phylogenetic tree was constructed using BEAST software using **(A)** GTR + C + I model from four full genome SNPs sequences. **(B)** Transversional substitution model (TVMef) from four full genome SNPs sequences.

8919, and 92,326, respectively. Most of the loci were unique for *E. granulosus* G1 (226,350) and *E. canadensis* G7 (335,728) rather than for *E. oligarthrus* (40,386). On the other hand, the number of shared loci containing the same nucleotide change among *E. canadensis* G7, *E. granulosus* G1, and *E. multilocularis* was 12,277 with respect to *E. oligarthrus* (*E. oligarthrus* used as reference). The number of shared loci between *E. multilocularis* and *E. granulosus* G1, between *E. multilocularis* and *E. canadensis*

G7, and between *E. canadensis* G7 and *E. granulosus* G1 was 8675, 8875, and 54,792 respectively. Furthermore, and similar to the above results, the highest numbers of unique loci were for *E. granulosus* G1 (133,417) and *E. canadensis* G7 (242,815), rather than for *E. multilocularis* (60,455) (**Figure 3**).

In order to gain accuracy and supporting evidence for our previous results, we reconstructed the *Echinococcus* phylogeny with nuclear molecular markers previously described (Kinkar

FIGURE 3 | Venn diagram of shared loci between the different *Echinococcus* species using *E. multilocularis* **(A)** and *E. oligarthrus* **(B)** as reference. The numbers under each species name indicate the number of SNPs in the species against the corresponding reference genome. The numbers in the overlap region indicate the number of SNPs with the same polymorphism at the same locus between the species analyzed.

et al., 2017). Nuclear molecular marker sequences of all the available *Echinococcus* species were concatenated and then were aligned and analyzed using both the Bayesian and maximum likelihood methods. This analysis involved the study of 1552 nucleotides that allowed to root the trees and reinforced the basal position of *E. oligarthrus* (**Supplementary Figure 5D**). The phylogenetic topology obtained was consistent with previous phylogenetic analyses of (Nakao et al., 2013b) that have placed *E. oligarthrus* (and *E. vogeli*) in a basal position for the genus *Echinococcus*. Similar phylogenetic topologies were obtained with the two methods employed here (**Supplementary Figure 5E**).

DISCUSSION

In this work, we sequenced the first sylvatic species of *Echinococcus*, *E. oligarthrus*, isolated from its natural host and performed comparative genomics between both domestic and sylvatic species. One of the limitations of obtaining complete genomes of wildlife parasites from natural infections resides on the difficulty of obtaining DNA samples free of the host material. In particular, for parasites whose development occurs in intimate contact with the host tissue, such is the case of *E. oligarthrus*. Here, we performed an extensive effort to identify the target sequences of the parasite and assemble the genome. Several steps were applied before the final assembly was obtained, and the final assembly was compared with the *Echinococcus* genomes obtained previously by us (Maldonado et al., 2017). Even though the quality of *E. oligarthrus* genome assembly was lower than other *Echinococcus* genomes, it was high enough to locate the genes and perform comparative genomic analysis.

Hereby, we used these data to unravel the phylogeny of this genus. In previous studies, we described the genetic variation among three *Echinococcus* species (*E. canadensis* G7, *E. granulosus* G1, and *E. multilocularis*) and assessed the distribution of SNPs in the whole genome as well as the effect and the type of SNPs in the coding regions. In this work, we added a new genome to the analysis of genetic variants and studied the SNP distribution in each one of their chromosomes. Regarding the genetic diversity among the *Echinococcus* species, we found that *E. canadensis* G7 and *E. granulosus* G1 contained almost 10 orders of magnitude more SNPs than between *E. oligarthrus* and *E. multilocularis*. The SNP distribution observed is similar to the distribution described in our previous research (Maldonado et al., 2017) where the genetic diversity within the *E. granulosus sensu lato* species was high. On the other hand, the genetic diversity between the sylvatic species *E. oligarthrus* and *E. multilocularis* is remarkably lower. Furthermore, the genetic variability showed to be unequal for different chromosomes. This fact was revealed by a higher SNP density in larger chromosomes than in the smallest ones. In previous studies, we have also reported a higher gene density in larger chromosomes (Maldonado et al., 2018). However, since most of the SNPs produce synonymous nucleotide changes, the amino acid sequences derived from these genes, even those located in the larger chromosomes, are not altered by the presence of such changes and presumably neither their function.

For several years, the mitochondrial sequences were employed to analyze the *Echinococcus* phylogeny (for a review, see Nakao

et al., 2013a). However, the construction of phylogenetic trees based only on mitochondrial DNA data may be biased because it is maternally inherited and, therefore, under particular evolutionary forces that may not represent the evolutionary history for each species (Lymbery, 2017). Indeed, it has been suggested that nuclear sequences should be used when evaluating the phylogenetic positions of new *Echinococcus* isolates (Saarma et al., 2009). Here, we performed maximum likelihood and Bayesian phylogenetic analyses using nuclear DNA sequences and compared the results implementing different models including the best-fitted evolutionary model predicted by PartitionFinder 2 (Lanfear et al., 2016). For this purpose, we used all the shared loci within SNPs that were identified among the *Echinococcus* genomes in both whole genome and coding regions and under the strict criteria of having more than 20× depth coverage in all the species. Hereby, and adding previously described nuclear molecular markers, which provides high accuracy to our results, the tree topology retrieved as the most frequent reconstruction placed *E. oligarthrus* in a basal position. Based on these analyses, we conclude that *E. oligarthrus* may be one of the basal species of the genus *Echinococcus*, together with *E. multilocularis*. These findings also agree with our previous studies (Maldonado et al., 2017) where we proposed a basal sylvatic species that could have accumulated mutations over time until a speciation phenomenon could have given rise to *E. granulosus* G1 and *E. canadensis* G7, which afterwards would have diverged, independently increasing the genetic diversity. The fact that *E. canadensis* G7 and *E. granulosus* G1 share more homozygous polymorphic loci with the same variant supports the hypothesis of a basal sylvatic species. However, since the number of homozygous polymorphic loci with the same variant shared among three of the species is almost equal (12,282 for *E. multilocularis* and 12,277 for *E. oligarthrus* as reference genomes), this result does not allow one to resolve whether *E. multilocularis*, *E. oligarthrus*, or other unknown ancestral related species is the ancestral species from which modern *Echinococcus* genomes could have arisen, and thus remains unclear. This hypothesis could be further probed with the complete genome analyses of more *Echinococcus* species, which would be really useful to describe the complete evolutionary history of these parasites. One of the most interesting implications of the nuclear phylogeny based on SNP analysis found in this work is the position of the *Echinococcus* sylvatic species as basal to the genus *Echinococcus*; in addition, it also demonstrates relevant genetic similarities between *E. multilocuaris* and *E. oligarthrus*. This evolutionary framework may enable data-driven investigation of morphological features and developmental evolutionary studies that would provide relevant information about the neotropical echinococcosis. The generation of genome datasets from additional cestode species would further improve these findings. Phylogenetic studies have allowed the development of hypotheses about the evolutionary history of several taxonomic groups from other perspectives. Such is the case of parasitic organisms and the phylogeography of their hosts that helps to interpret parasite evolution in relation to the migratory patterns of their hosts and vice versa. In the neotropical region, several felids serve as definitive hosts for *E. oligarthrus*. Recently, we determined for the first time the presence of *E. oligarthrus* in ocelot (*Leopardus*

pardalis) and the puma (*Puma concolor*) in the north of Argentina using nuclear and mitochondrial molecular markers (Arrabal et al., 2017). Indeed, all the cases reported so far have come from South America (D'Alessandro and Rausch, 2008). In terms of phylogeography, the most suitable explanation is that carnivores originate from immigrants from North America and the ancestral species of *Echinococcus* migrated to South America together with their felid hosts. Early studies suggested that the differentiation of the species of *Leopardus* was likely facilitated by the formation of the Panamanian land bridge. The current hypothesis about felid evolution based on molecular phylogenetic studies suggests that the endemic neotropical felids (genus *Leopardus*) have diverged from other main felid lineages and that before the emergence of the Panamanian isthmus they could have migrated to South America. (Johnson et al., 2006). The finding of an archaic lineage of *Trichinella* in South America also supports the hypothesis of this early carnivore expansion (Pozio et al., 2009). Although originating in North America, the Puma currently has an extensive geographic range in South America and could explain the presence of *E. oligarthrus* in several felid species from the neotropical region. Regarding intermediate hosts, the rodents of the *Hystricomorpha* suborder, natural hosts of *E. oligarthrus*, are known to be the dominant small terrestrial herbivores in South America by the Miocene (Eisenberg et al., 1989). Hence, both the neotropical *Echinococcus* species and their respective hosts seem to have an ancient origin (Nakao et al., 2013b). The need for more genomes and analyses of host–parasite interactions are evident in order to further understand the co-evolution between this parasite and its felid host and the lifestyles of *Echinococcus* species.

Before sequencing the *E. oligarthrus* genome, there was a paucity of related molecular data of this organism. Indeed, only 47 nucleotide sequences have been reported so far, representing only nine genes. Most of them were obtained with the sole purpose of being used as molecular markers (e.g., 18s rRNA nuclear gene or cox1 mitochondrial gene). The scarce sequences information about this parasite was significantly improved through our effort to sequence, assemble, and annotate the genome of *E. oligarthrus*. Despite the fragmented nature of the assembly, we have thoroughly analyzed the gene content in comparisons with other members of the genus. Most of the core genes are identifiable in the genome even though many genes are present as fragments within the assembly. Although many genes presented coverage < 50%, the GO terms distribution found for *E. oligarthrus* was according to that observed in other *Echinococcus* species. Even more, we found that many genes are typically conserved in all the cestodes species. Indeed, these genes composed the core of cestodes genes according to Maldonado et al. (2017), which means that these genes can be found in all the cestodes species whose genomes have been sequenced so far. Therefore, both the genome and the gene annotations of *E. oligarthrus* are suitable to be used in several bioinformatic and comparative analysis as well as to guide hummed and molecular assays. Several works have reported the molecular and cellular mechanisms implicated in the larvae development of *E. multilocularis* and *E. granulosus sensu stricto* (Brehm and Koziol, 2017) but not much is known about the neotropical *Echinococcus* species that have a particular larvae morphology (D'Alessandro and Rausch, 2008). By means of this new genome, we expanded the repertoire of the available genes and reached the 85% of the genes expected to be present in *Echinococcus*, providing for the first time a large set of proteins of a neotropical *Echinococcus* species that can be further studied. Until now, only two families of coding genes implicated in host–parasite interactions had been sequenced from *E. oligarthrus*, the antigen Eg95 and the partial sequence of the antigen B (Haag et al., 2006, Haag et al., 2009). In this regard, we searched for proteins already known to be involved in host–parasite interactions, including genes that are responsible for larval development and signaling pathways in *Echinococcus*. Genes such as Wnt or TNFα receptor and putative regulatory genomic sequences of thousands of important genes were described here. By means of this work, the study of the genes and genomic regulatory regions of the neotropical species *E. oligarthrus* are now reachable by the scientific community. The data obtained here will allow the design of data-driven experiments of gene expression that will provide clues about the particular behavior of the parasite into the mammalian hosts and its differences between sylvatic and domestic species.

DATA AVAILABILITY

The assembled sequences of the E. oligarthrus genome were deposited in ENA (BioProject PRJEB31222, http://www.ebi.ac.uk/ena/data/view/PRJEB31222). The sequences and the annotation data can also be downloaded from our FlatDB project web page (http://www.bmhid.org/flatdb/). All the data generated and analyzed in this study are included in this published article and within the **Supplementary Material**.

AUTHOR CONTRIBUTIONS

LM performed the bioinformatics analysis and wrote the manuscript. JA collected the parasite material and performed the genetic and morphological analysis. LM and LK designed the study. GO constructed the libraries and performed the sequencing. LM, LK, MR, and GO wrote and revised the manuscript. All authors read and approved the manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00708/full#supplementary-material

REFERENCES

Arrabal, J. P., Avila, H. G., Rivero, M. R., Camicia, F., Salas, M. M., Costa, S. A., et al. (2017). *Echinococcus oligarthrus* in the subtropical region of Argentina: first integration of morphological and molecular analyses determines two distinct populations. *Vet. Parasitol.* 240, 60–67. doi: 10.1016/j.vetpar.2017.03.019

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene Ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Barnett, D. W., Garrison, E. K., Quinlan, A. R., Strömberg, M. P., and Marth, G. T. (2011). BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* 27, 1691–1692. doi: 10.1093/bioinformatics/btr174

Birney, E., Clamp, M., and Durbin, R. (2004). GeneWise and genomewise. *Genome Res.* 14, 988. doi: 10.1101/gr.1865504

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Brehm, K., and Koziol, U. (2017). *Echinococcus*–host interactions at cellular and molecular levels. *Adv. Parasitol.* 95, 147–212. doi: 10.1016/bs.apar.2016.09.001

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin).* 6, 80–92. doi: 10.4161/fly.19695

D'Alessandro, A., and Rausch, R. L. (2008). New aspects of neotropical polycystic (*Echinococcus vogeli*) and unicystic (*Echinococcus oligarthrus*) echinococcosis. *Clin. Microbiol. Rev.* 21, 380–401. doi: 10.1128/CMR.00050-07

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330

Danecek, P., and McCarthy, S. A. (2017). BCFtools/csq: haplotype-aware variant consequences. *Bioinformatics* 33, 2037–2039. doi: 10.1093/bioinformatics/btx100

De Baets, K., Dentzien-Dias, P., Upeniece, I., Verneau, O., and Donoghue, P. C. J. (2015). Constraining the deep origin of parasitic flatworms and host-interactions with fossil evidence. *Adv. Parasitol.*, 90, 93–135. doi: 10.1016/bs.apar.2015.06.002

Eisenberg, J. F., Reid, F., Redford, K. H., and Bonner, S. J., (1989). *Mammals of the Neotropics*. University of Chicago Press, Chicago.

Finn, R. D. (2006). Pfam: Clans, web tools and services. *Nucleic Acids Res.* 34, D247–D251. doi: 10.1093/nar/gkj149

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086

Haag, K. L., Gottstein, B., and Ayala, F. J. (2009). The EG95 antigen of *Echinococcus* spp. contains positively selected amino acids, which may influence host specificity and vaccine efficacy. *PLoS One* e5362, 4. doi: 10.1371/journal.pone.0005362

Haag, K. L., Zanotto, P. M. A., Alves-Junior, L., Gasser, R. B., Zaha, A., and Ayala, F. J. (2006). Searching for antigen B genes and their adaptive sites in distinct strains and species of the helminth *Echinococcus. Infect. Genet. Evol.* 6, 251–261. doi: 10.1016/j.meegid.2005.07.003

Hahn, C., Fromm, B., and Bachmann, L. (2014). Comparative genomics of flatworms (Platyhelminthes) reveals shared genomic features of ecto- and endoparasitic neodermata. *Genome Biol. Evol.* 6, 1105–1117. doi: 10.1093/gbe/evu078

Hüttner, M., Nakao, M., Wassermann, T., Siefert, L., Boomker, J. D. F., Dinkel, A., et al. (2008). Genetic characterization and phylogenetic position of *Echinococcus felidis* Ortlepp, 1937 (Cestoda: Taeniidae) from the African lion. *Int. J. Parasitol.* 38, 861–868. doi: 10.1016/j.ijpara.2007.10.013

Johnson, W. E., Eizirik, E., Pecon-Slattery, J., Murphy, W. J., Antunes, A., Teeling, E., et al. (2006). The late Miocene radiation of modern Felidae: a genetic assessment. *Science* 311, 73–77. doi: 10.1126/science.1122277

Keibler, E., and Brent, M. R. (2003). Eval: a software package for analysis of genome annotations. *BMC Bioinform.* 4, 50. doi: 10.1186/1471-2105-4-50

Kinkar, L., Laurimäe, T., Sharbatkhori, M., Mirhendi, H., Kia, E. B., Ponce-Gordo, F., et al. (2017). New mitogenome and nuclear evidence on the phylogeny and taxonomy of the highly zoonotic tapeworm *Echinococcus granulosus* sensu stricto. *Infect. Genet. Evol.* 52, 52–58. doi: 10.1016/j.meegid.2017.04.023

Knapp, J., Nakao, M., Yanagida, T., Okamoto, M., Saarma, U., Lavikainen, A., et al. (2011). Phylogenetic relationships within *Echinococcus* and *Taenia* tapeworms (Cestoda: Taeniidae): an inference from nuclear protein-coding genes. *Mol. Phylogenet. Evol.* 61, 628–638. doi: 10.1016/j.ympev.2011.07.022

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Laetsch, D. R., and Blaxter, M. L. (2017). BlobTools: interrogation of genome assemblies. *F1000Research* 6, 1287. doi: 10.12688/f1000research.12232.1

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., and Calcott, B. (2016). PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* 34, msw260. doi: 10.1093/molbev/msw260

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Lymbery, A. J. (2017). Phylogenetic pattern, evolutionary processes and species delimitation in the genus *Echinococcus. Adv. Parasitol.*, 95, 111–145. doi: 10.1016/bs.apar.2016.07.002

Lymbery, A. J., Jenkins, E. J., Schurer, J. M., and Thompson, R. C. A. (2015). *Echinococcus canadensis, E. borealis*, and *E. intermedius*. What's in a name? *Trends Parasitol.* 31, 23–29. doi: 10.1016/j.pt.2014.11.003

Maldonado, L. L., Assis, J., Araújo, F. M. G., Salim, A. C. M., Macchiaroli, N., Cucher, M., et al. (2017). The *Echinococcus canadensis* (G7) genome: a key knowledge of parasitic platyhelminth human diseases. *BMC Genomics* 18, 204. doi: 10.1186/s12864-017-3574-0

Maldonado, L. L., Stegmayer, G., Milone, D. H., Oliveira, G., Rosenzvit, M., and Kamenetzky, L. (2018). Whole genome analysis of codon usage in. *Echinococcus. Mol. Biochem. Parasitol.* 225, 54–56.doi: 10.1016/j.molbiopara.2018.08.001

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110

McManus, D. P. (2013). Current status of the genetics and molecular taxonomy of *Echinococcus* species. *Parasitology* 140, 1617–1623. doi: 10.1017/S0031182013000802

Nakao, M., Lavikainen, A., Yanagida, T., and Ito, A. (2013a). Phylogenetic systematics of the genus *Echinococcus* (Cestoda: Taeniidae). *Int. J. Parasitol.* 43, 1017–1029. doi: 10.1016/j.ijpara.2013.06.002

Nakao, M., McManus, D. P., Schantz, P. M., Craig, P. S., and Ito, A. (2007). A molecular phylogeny of the genus *Echinococcus* inferred from complete mitochondrial genomes. *Parasitology* 134, 713–722. doi: 10.1017/S0031182006001934

Nakao, M., Yanagida, T., Konyaev, S., Lavikainen, A., Odnokurtsev, V. A., Zaikov, V. A., et al. (2013b). Mitochondrial phylogeny of the genus *Echinococcus* (Cestoda: Taeniidae) with emphasis on relationships among *Echinococcus canadensis* genotypes. *Parasitology* 140, 1625–1636. doi: 10.1017/S0031182013000565

Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C., and Durbin, R. (2016). BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics* 32, 1749–1751. doi: 10.1093/bioinformatics/btw044

Pozio, E., Hoberg, E., La Rosa, G., and Zarlenga, D. S. (2009). Molecular taxonomy, phylogeny and biogeography of nematodes belonging to the *Trichinella* genus. *Infect. Genet. Evol.* 9, 606–616. doi: 10.1016/j.meegid.2009.03.003

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., et al. (2005). InterProScan: protein domains identifier. *Nucleic Acids Res.* 33, W116–W120. doi: 10.1093/nar/gki442

Quinlan, A. R. (2014). BEDTools: the Swiss-Army tool for genome feature analysis. *Curr. Protoc. Bioinforma.* 47, 11.12.1–34. doi: 10.1002/0471250953.bi1112s47

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033

Rausch, R. L., Journal, T., and Jun, N. (1967). A consideration of infraspecific categories in the genus *Echinococcus* Rudolphi, 1801 (Cestoda : Taeniidae). *Genus* 53, 484–491. doi: 10.2307/3276706

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* 16, 276–277. doi: 10.1016/S0168-9525(00)02024-2

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029

Saarma, U., Jõgisalu, I., Moks, E., Varcasia, A., Lavikainen, A., Oksanen, A., et al. (2009). A novel phylogeny for the genus *Echinococcus*, based on nuclear data, challenges relationships based on mitochondrial evidence. *Parasitology* 136, 317–328. doi: 10.1017/S0031182008005453

Safonova, Y., Bankevich, A., and Pevzner, P. A. (2014). DIPSPADES: Assembler for highly polymorphic diploid genomes. *J Comput Biol.* 22(6): 528–545. doi: 10.1007/978-3-319-05269-4_21

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351

Slater, G., Birney, E., Box, G., Smith, T., Waterman, M., Altschul, S., et al. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinform.* 6, 31. doi: 10.1186/1471-2105-6-31

Soares M. do, C. P., Rodrigues, A. L. dos S., Moreira Silva, C. A., Brito, E. M. de F., Gomes-Gouvêa, M. S., Corrêa, I. R. dos S., et al. (2013). Anatomo-clinical and molecular description of liver neotropical echinococcosis caused by *Echinococcus oligarthrus* in human host. *Acta Trop.* 125, 110–114. doi: 10.1016/j.actatropica.2012.09.004

Stemmer, K., Burghardt, I., Mayer, C., Reinicke, G. B., Wägele, H., et al. (2013). Morphological and genetic analyses of xeniid soft coral diversity (Octocorallia; Alcyonacea). *Org. Divers. Evol.* 13, 135–150. doi: 10.1007/s13127-012-0119-x

Tamazian, G., Dobrynin, P., Krasheninnikova, K., Komissarov, A., Koepfli, K.-P., and O'Brien, S. J. (2016). Chromosomer: a reference-based genome arrangement tool for producing draft chromosome sequences. *Gigascience* 5, 38. doi: 10.1186/s13742-016-0141-6

Tsai, I. J., Zarowiecki, M., Holroyd, N., Garciarrubio, A., Sanchez-Flores, A., Brooks, K. L., et al. (2013). The genomes of four tapeworm species reveal adaptations to parasitism. *Nature* 496, 57–63. doi: 10.1038/nature12031

Verster, A. (1969). A taxonomic revision of the genus Taenia Linnaeus, 1758 s. str. *J. Vet. Res.* 36 (1), 3-58. http://hdl.handle.net/2263/53847.

Xiao, N., Qiu, J., Nakao, M., Li, T., Yang, W., Chen, X., et al. (2006). *Echinococcus shiquicus*, a new species from the Qinghai–Tibet plateau region of China: Discovery and epidemiological implications. *Parasit. Int.* 55, S233–S236. doi: 10.1016/j.parint.2005.11.035

Zheng, H., Zhang, W., Zhang, L., Zhang, Z., Li, J., Lu, G., et al. (2013). The genome of the hydatid tapeworm *Echinococcus granulosus*. *Nat. Genet.* 45, 1168–1175. doi: 10.1038/ng.2757

Compositional Analysis of Flatworm Genomes Shows Strong Codon Usage Biases Across All Classes

Guillermo Lamolle[1], Santiago Fontenla[1], Gastón Rijo[1], Jose F. Tort[1]*
and Pablo Smircich[2,3]*

[1] Departamento de Genética, Facultad de Medicina, Universidad de la Republica, UDELAR, Montevideo, Uruguay,
[2] Departamento de Genómica, Instituto de Investigaciones Biológicas Clemente Estable, IIBCE, MEC, Montevideo,
Uruguay, [3] Laboratorio de Interacciones Moleculares, Facultad de Ciencias, Universidad de la Republica, UDELAR,
Montevideo, Uruguay

In the present work, we performed a comparative genome-wide analysis of 22 species representative of the main clades and lifestyles of the phylum Platyhelminthes. We selected a set of 700 orthologous genes conserved in all species, measuring changes in GC content, codon, and amino acid usage in orthologous positions. Values of 3rd codon position GC spanned over a wide range, allowing to discriminate two distinctive clusters within freshwater turbellarians, Cestodes and Trematodes respectively. Furthermore, a hierarchical clustering of codon usage data differs remarkably from the phylogenetic tree. Additionally, we detected a synonymous codon usage bias that was more dramatic in extreme GC-poor or GC-rich genomes, i.e., GC-poor Schistosomes preferred to use AT-rich terminated synonymous codons, while GC-rich *M. lignano* showed the opposite behavior. Interestingly, these biases impacted the amino acidic usage, with preferred amino acids encoded by codons following the GC content trend. These are associated with non-synonymous substitutions at orthologous positions. The detailed analysis of the synonymous and non-synonymous changes provides evidence for a two-hit mechanism where both mutation and selection forces drive the diverse coding strategies of flatworms.

Keywords: flatworms, GC content, synonymous codons, codon usage, non-synonymous substitutions, amino acid usage, mutation, selection

INTRODUCTION

The phylum Platyhelminthes with more than 30,000 species is one of the major phyla of invertebrate animals containing an enormous diversity of life forms that had colonized very diverse niches (Caira and Littlewood, 2013). Almost three quarters of the flatworms are parasitic and belong to the Neodermata, a monophyletic clade characterized by a syncytial tegument and the presence of diverse specialized organs to attach to hosts like suckers and hooks. The Neodermata comprise three classes: the Monogenea (primarily external parasites of cold-blooded aquatic vertebrates), the Cestoda (obligate endoparasites of vertebrates), and the Trematoda (endoparasites of vertebrates as adults, with intermediate stages endoparasitic in other invertebrates, mainly mollusks) (Caira and Littlewood, 2013). Besides the parasitic Neodermatans, an enormous diversity of species occurs in seas, rivers, and lakes and on all continental land masses comprising one of the successful phyla of invertebrates (Collins, 2017). A few species exist as either commensals or occasional parasites of

invertebrates, but most of them are free-living predator species forming a single paraphyletic group collectively referred to as "turbellarians" (Caira and Littlewood, 2013). Studies based on rRNA (Larsson and Jondelius, 2008; Laumer and Giribet, 2014) and transcriptomic data (Egger et al., 2015; Laumer et al., 2015) showed that the phylum Platyhelminthes split early into two clades: the ancestral Catenulida and the Rhabditophora, which includes several free-living orders and the Neodermatans (**Table 1**). Taxonomically, the Macrostomorpha was placed as the earliest diverging Rhabditophoran lineage and the order Tricladida, which contains the model organism *Schmidtea mediterranea*, as part of the later-evolved "turbellarians."

The huge diversity of flatworm's life forms seems to be paralleled at genomic level. The recent publication of several genomic assemblies of the phylum (most of them corresponding to parasitic Neodermatans) has revealed a wide genomic diversity. For example, genome sizes range from 67 or 104 Mbases in the monogenean *Gyrodactylus salaris* or the cestode *Hydatigera taeniaeformis*, respectively, to 1,200 Mbases in the trematode *Fasciola hepatica* (Coghlan et al., 2019). Interestingly, this variation has little correlation with gene set completeness among genomes and is mostly due to non-coding elements, including repetitive and non-repetitive elements, with repeat content ranging from less than 4% in the smallest genomes of cestodes to 68% in *Fasciola hepatica*. Additionally, guanine and cytosine (GC) contents are very diverse from 28% in the planaria *S. mediterranea* and 33% in Monogenea *Gyrodactylus salaris*, to more than 45% in *M. lignano* and the food-borne trematodes (FBT) *F. hepatica*, *C. sinensis,* and *O. volvulus* (Coghlan et al., 2019).

We wondered if these large variations in genomic composition and structure could be correlated with the morphological and ecological diversity. It is well known that genomic GC content determines codon usage across species (Bernardi and Bernardi, 1985; Plotkin and Kudla, 2011) and the use of alternative synonymous codons is a non-random process (Sharp et al., 2010; Plotkin and Kudla, 2011). Due to the degeneracy of the genetic code, most amino acids, with the exceptions of methionine and tryptophan, are encoded by more than one codon. Codon usage bias (CUB) is a phenomenon where synonymous codons are not used with equal frequencies in coding DNA. It has been suggested that codon usage bias is the result of an equilibrium between mutational bias and natural selection and that natural selection could be acting in presumably highly expressed genes (Sharp et al., 2010; Plotkin and Kudla, 2011). Besides the effect at synonymous codon usage, it has been shown that strong GC bias could lead to changes in amino acid frequencies (Behura and Severson, 2013; Li et al., 2015). While this has not been explored widely in flatworms, several advances have been made in nematodes (Cutter et al., 2006; Mitreva et al., 2006; Mazumder et al., 2017a; Mazumder et al., 2017b). It is not clear yet how genomic GC differences could be influencing the codon usage and amino acid composition of proteins in Platyhelminthes and if these variations correlate with the ecological and physiological diversity in the phylum.

First reports of flatworm codon usage predated the genomic era and were based on a low representative number of sequences in Schistosomes and *Echinococcus*. Heterogeneity was evidenced since Schistosomes preferred A+T-rich codons, while *Echinococcus* favored GC3-rich codons (Meadows and Simpson, 1989; Alvarez et al., 1993; Kalinna and McManus, 1994; Milho and Tracy, 1995). Further analysis in larger sets of genes showed that codon bias was not uniformly distributed between genes introducing the possibility of isochores (regions that differ in GC content) in the genomes of flatworms (Ellis and Morrison, 1995; Ellis et al., 1995). In agreement, a more recent compositional analysis of

TABLE 1 | List of species analyzed and their GC content. FL, Free-living; PR, parasitic; G.GC, genomic GC percentage; T.GC, transcript (CDS) GC percentage.

Class	Subclass/order	Species	Abbreviation	Habitat	G.GC	T.GC
CATENULIDA		*Stenostomum leucops*	Sleu	FL		54.0
RHABDITOPHORA	Macrostomorpha	*Macrostomum lignano*	Mlig	FL	45.9	58.7
RHABDITOPHORA	Lecithoepitheliata	*Geocentrophora applanata*	Gapp	FL		37.8
RHABDITOPHORA	Polycladida	*Prosthecereaus vittatus*	Pvit	FL		46.4
RHABDITOPHORA	Neodalyellida/Rhabdocoela	*Rhynchomesostoma rostratum*	Rros	FL		40.1
RHABDITOPHORA	Dalyellioida/Fecampiida	*Kronborgia amphipodicola*	Kamp	PR		40.3
RHABDITOPHORA	Seriata/Proseriata	*Monocelis fusca*	Mfus	FL		40.8
RHABDITOPHORA	Seriata/Bothrioplanida	*Botrhioplana semperi*	Bsem	FL		53.0
RHABDITOPHORA	Seriata/Tricladida	*Schmidtea mediterranea*	Smed	FL	29.9	36.6
MONOGENEA	Monopisthocotylea/Gyrodactylidea	*Gyrodactilus salaris*	Gsal	PR	33.9	43.3
MONOGENEA	Polyopisthocotylea/Polystomatidea	*Protopolystoma xenopodis*	Pxen	PR	37.7	50.2
TREMATODA	Digenea/Strigeidida	*Schistosoma mansoni*	Sman	PR	35.5	36.0
TREMATODA	Digenea/Strigeidida	*Schistosoma japonicum*	Sjap	PR	34.1	36.0
TREMATODA	Digenea/Strigeidida	*Trichobilharzia regenti*	Treg	PR	37.4	37.2
TREMATODA	Digenea/Plagiorchiida	*Fasciola hepatica*	Fhep	PR	44.1	47.8
TREMATODA	Digenea/Opisthorchiida	*Clonorchis sinensis*	Csin	PR	44.0	48.4
TREMATODA	Digenea/Opisthorchiida	*Opistorchis viverrini*	Oviv	PR	43.8	48.5
CESTODA	Eucestoda/Diphyllobothriidea	*Schistocephalus solidus*	Ssol	PR	43.0	51.9
CESTODA	Eucestoda/Cyclophyllidea	*Mesocestoides corti*	Mcor	PR	36.7	51.4
CESTODA	Eucestoda/Cyclophyllidea	*Hymenolepis diminuta*	Hdim	PR	35.2	44.2
CESTODA	Eucestoda/Cyclophyllidea	*Echinococcus granulosus*	Egra	PR	41.9	50.0
CESTODA	Eucestoda/Cyclophyllidea	*Echinococcus multilocularis*	Emul	PR	42.2	49.9

FL, free-living; PR, parasitic; G.GC, genomic GC percentage; T.GC, transcript (CDS) GC percentage.

the *S. mansoni* genome reported an isochore-like organization (Lamolle et al., 2016). Early studies analyzing the forces behind codon bias found evidences of both mutational pressure (Musto et al., 1998) in *S. mansoni* and selection (Fernandez et al., 2001) in *Echinococcus* spp. as preponderant forces. More recently, studies on *S. haematobium* and *S. japonicum* confirmed a major role of natural selection in shaping the codon usage bias in these species (Mazumder et al., 2017a; Mazumder et al., 2017b). Several studies analyzed codon usage on the available genomes and transcriptomes of *Taenidae* species showing weak codon bias and a higher GC3 in highly expressed genes explained by combined mutational and selection forces (Chen et al., 2013; Yang et al., 2014; Yang et al., 2015; Huang et al., 2017). A more preponderant contribution of selection shaping codon usage was identified in a comparative analysis in *Echinococcus* species (Maldonado et al., 2018). While these studies highlight that platyhelminthes are compositionally varied, they are focused just in the schistosomes and tapeworms. We took advantage of the wide array of transcriptomes and genomes now available to extend the study to a phylum-wide analysis of codon usage patterns, as a proxy of the molecular organization of flatworm genomes. We performed a comparative analysis at the genomic level of 22 species representative of the main clades and lifestyles of the phylum Platyhelminthes. Within these species, we picked a set of 700 orthologous gene groups conserved across the 22 species and measured changes in GC content, codon, and amino acid usage in orthologous positions. We found a class independent-wide diversity in codon and amino acid usages. Based on the study of orthologous positions in selected pairs of species with diverse GC content, we provide evidence of a combined contribution of mutational forces and selection that enforced synonymous codon usage bias and differential amino acid usage.

METHODS

Data Acquisition
Genomic and coding sequences of 22 flatworm species were used in this work. To ease data visualization, a four-letter code was used to name the species (**Table 1**). Genomic and transcriptomic data of Mlig, Smed, Gsal, Pxen, Csin, Oviv, Fhep, Treg, Sjap, Sman, Mcor, Hdim, Egra, Emul, and Ssol were obtained from the public repository Wormbase parasite (Howe et al., 2017) (https://parasite.wormbase.org/). Transcriptomic data on Sleu, Gapp, Pvit, Rros, Mfus, Kamp, and Bsem were generated by Laumer et al. (2015) and downloaded from the public repository Data Dryad (doi:10.5061/dryad.622q4).

Orthologues Determination
In-house Perl and Bash scripts that implemented a BLASTp best reciprocal hit strategy were used to identify a core of orthologous genes. An e-value cutoff of 1e-5 was used to define significant hits. The restrictive method produced one orthologue gene per species. A total of 700 orthologous groups were detected in all 22 species, and these sequences were used for the analysis, adding to more than 8 million codons analyzed (8.242.428).

Expression data for *S. mansoni* in reads per kilobase per million (RPKM) were taken from the study of Protasio et al. (2012). Expression data were available for 696 of the 700 *S. mansoni* orthologs. Expression data for the adult stages of *F. hepatica*, *E. granulosus*, *H. diminuta*, *S. mediterranea*, and *M. lignano* were downloaded from WormBase Parasite (Howe et al., 2017).

Gene Alignment and Phylogenetic Tree
Each group of 22 orthologous sequences were translated with an in-house Perl script and aligned individually with Mafft (Katoh and Standley, 2013). Individual alignments were concatenated into a unique alignment. This alignment was used to build a phylogenetic tree with PhyML (Guindon et al., 2010). PhyML was run with the following options: -b -4, to calculate statistical branch support, -s BEST, for tree topology estimation, -m LG, to indicate the model substitution matrix, and -o tl, for tree topology and branch length optimization.

For the hierarchical clustering based on GC content of synonymous codons (RSCU), several clusters were built using hclust from R Stats package with different option settings (R Core Team, 2019). A final consensus cluster was made with the Ape package (Paradis et al., 2004), which retained the most frequent groupings.

Codon Usage Analysis
Codon usage and compositional analyzes were done in R with the package seqinR (Charif and Lobry, 2007). Correlations between frequencies of each codon and GC3 were represented as heatmap with the R "Corrplot" package (Wei and Simko, 2017). In-house R scripts were used to evaluate significance of changes in frequencies between high- and low-expressed gene sets and defined preferred codons. A codon was considered "preferred" if its frequency (RSCU) significantly increases in a set of high-expression genes, compared with a low-expression set, regardless of whether it becomes the main codon for that amino acid or not. Correspondence analysis (COA) was performed in R.

Neutrality and Effective Number of Codons Plots
Neutrality plots (Sueoka, 1988) (GC3 vs GC12) of the 22 species were used to evaluate the relationship among the three codon positions. Additionally, a unique plot showing general GC3–GC12 for all species was calculated by using a concatenated super gene for each species.

The effective number of codons (ENC) is used to quantify the variation in codon usage, ranging from 20 (when only one codon per amino acid is used) to 61 (when all possible codons are used). GC3 vs ENC charts are useful to estimate selection contribution to CUB. Expected values of ENC based on mutation pressure generate a bell curve, so in these charts, the points that fall directly on the curve represent genes with neutral evolution, while the points under the curve suggest action of natural selection (Wright, 1990).

Amino Acid and Codon Substitutions Matrices

From the amino acid alignments of each COG, the sites that had gaps in one or more sequences were eliminated. Degapped COGs with less than 35 amino acids were eliminated. The resulting sequences were then concatenated, generating a "super-peptide" (without gaps) for each species. Then, the amino acid changes between each pair of species were counted (with a homemade R script), creating 20 by 20 substitution matrices. Each value of the matrix ($A_{Z,X}$) represents how many times amino acid Z is present in one species, while amino acid X is present in the corresponding orthologous position in the other species, being, therefore, an asymmetric matrix. The diagonal of the matrix represents the unchanged sites, while the sum of the remaining values in each column or row represents the total substitutions for each amino acid. To test for deviations in the amino acid usage between species, the total count for each amino acid in the species of a pair was calculated, and the average was considered as expected value to perform chi-square tests (**Figure 5**). For each reciprocal changes in the matrix ($A_{Z,X}$, $B_{X,Z}$), a chi-square test was performed considering the average of the counts as expected value. For simplicity of analysis, we focused in three comparisons between species with different global GC: cestodes (Hdim and Egra), trematodes (Sman and Fhep), and free-living species (Smed and Mlig). The last comparison involved the two more divergent species in GC content.

Based on the back-translation of the alignments, we generated a 61 × 61 (stop codons deleted) codon substitution matrix for the six selected species following similar procedures as the ones described in the previous section.

RESULTS

Global GC Composition Varies Across Diverse Flatworm Taxa

As a first approach to analyze if there is a GC compositional difference in the phylum Platyhelminthes, we inspected the difference in the global genomic and transcriptomic G+C content. At first glance, it was clear that there is no correlation between genomic and transcriptomic GC, so it was not possible to use transcript GC to infer genomic GC. In most of the species, transcripts were GC richer than global genomic GC with the only exception of *T. regenti* (**Table 1**). However, while Schistosomatidae species show almost no difference in GC content between the overall genome and the coding region, Cestodes transcripts, for example, were on average 9.7% GC richer than all the genome considered together.
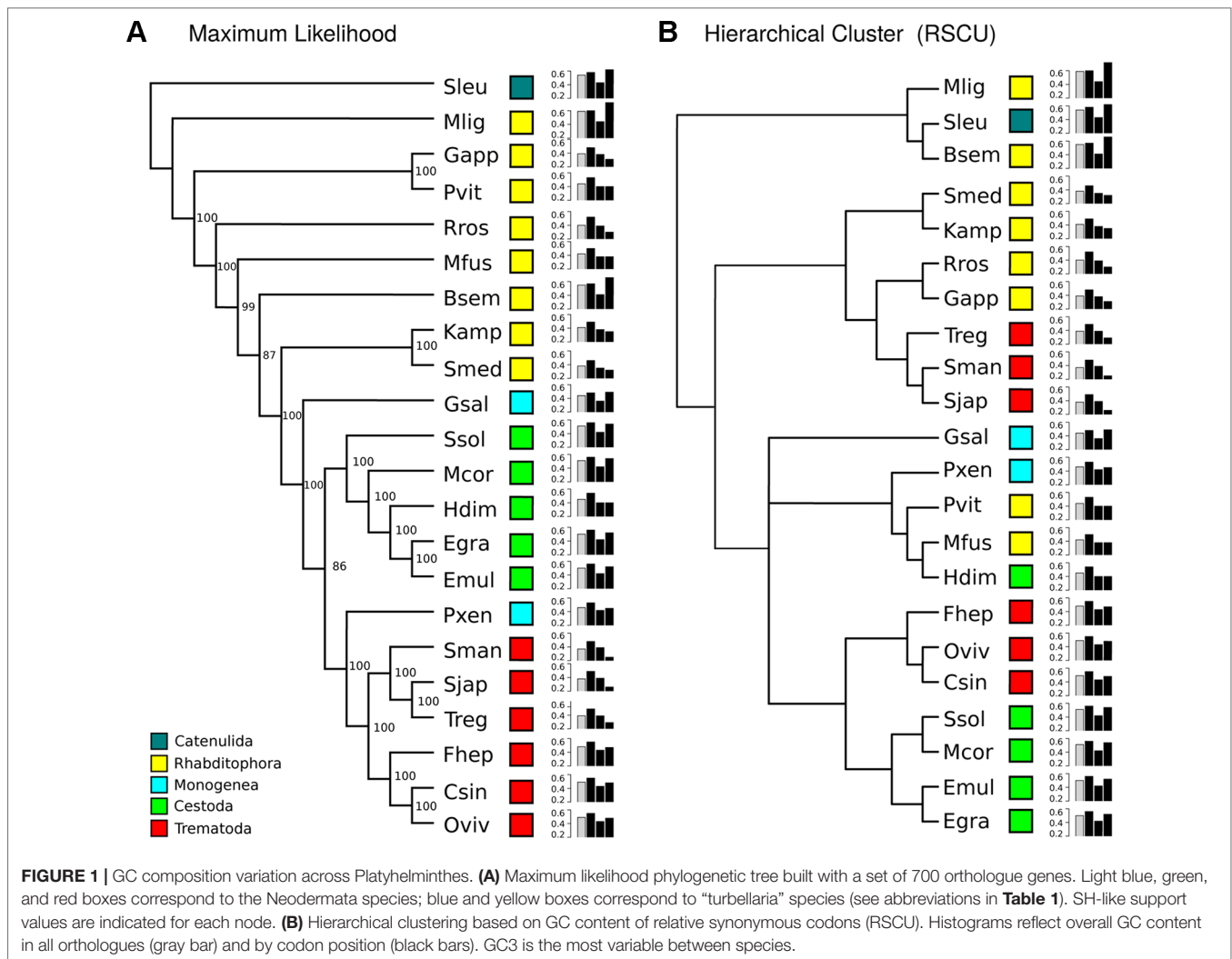
GC Composition Varies Across Diverse Flatworm Taxa

To further analyze the GC composition in the coding region, we searched for a set of orthologous conserved genes in the available genomes and transcriptomes. Based on a best reciprocal hit BLAST search, we selected 700 orthologous genes present in the 22 species. A maximum likelihood tree confirmed that the orthologous groups strongly represented the accepted phylogeny of the species analyzed (organisms form the same group cluster together, with the only exception of the two Monogenea, which represent two distinct subclasses) (**Figure 1A**). Since GC varies across the different species, we calculated the relative synonymous codon usage (RSCU) and the GC values by codon position in this set of conserved genes. The clustering of the species based on the relative synonymous codon usage (RSCU) data showed an important reorganization respect to the phylogenetic tree (**Figure 1B**). Three main clusters were clearly appreciable: the first with high GC3 (with free-living species), the second with low GC3 values (including other free-living species and the schistosomatids), and a third with intermediate GC3 values. GC2 was the less variable between groups (0.42, 0.38, and 0.41 on average in groups 1, 2, and 3, respectively). Additionally, we noticed three subgroups within group 3: one that had lower GC1–2 than the rest but had high GC3 composed only by the monogean *G. salaris*; the cluster of *P. xenopodis*, *P. vittatus*, *M. fusca*, and the cestode *H. diminuta* that had lower GC3; and the subgroup composed by trematodes (*F. hepatica*, *O. viverrini*, *C. sinensis*) and cestodes (*S. solidus*, *M. corti*, *E. multilocularis*, and *E. granulosus*) that had higher G1-3 compared with other species of the group. This shows that global synonymous codon usage varies widely across the phylum.

The relation of GC values in 1st and 2nd position versus those presented in 3rd codon position (neutrality plot) is usually used to evaluate if the variations in codon usage are driven by mutation or selection. Neutrality plots for the 22 organisms based on the 700 orthologue genes were analyzed. In all cases, low slopes were found for the regression curve (maximum value of 0.2) (**Supplementary Figure 1**). Careful inspection of the plots indicates that this can be explained by a low variability of GC1/2 among the genes (ranges between 0.4 and 0.6). These results suggest a contribution of selection in shaping codon usage for these organisms. To visualize all species together, we plotted the GC1–2 versus GC3 of the concatenated orthologue groups (COGs) for each species (see Methods). Expectedly, while GC1–2 presented little variation across species, the best discriminator was variation at GC3 (**Figure 2**). For example, between the most GC biased genomes, the GC-poor *S. mediterranea* and the GC-rich *M. lignano*, there was only 10% variation in GC1–2 axis but 40% variation in GC3. Interestingly, based on GC3 variability, we found two distinctive clusters within the freshwater "turbellarians," trematodes and cestodes. The trematodes species clearly differentiated the blood-dwelling flukes grouped on the lowest side of the GC3 spectrum to the food-borne liver flukes allocated in the middle upper GC3 range. Similarly, cestodes tend to cluster on the upper side of the GC3 range with the exception of the Hymenolepidae that fall on the lower-middle of the GC range. Freshwater "turbellarians" showed the largest variability in both the GC12 and GC3 range grouping into very distant clusters. However, we found no clear evolutionary-GC content correlation as species belonging to different lineages were mixed in both groups.

FIGURE 1 | GC composition variation across Platyhelminthes. **(A)** Maximum likelihood phylogenetic tree built with a set of 700 orthologue genes. Light blue, green, and red boxes correspond to the Neodermata species; blue and yellow boxes correspond to "turbellaria" species (see abbreviations in **Table 1**). SH-like support values are indicated for each node. **(B)** Hierarchical clustering based on GC content of relative synonymous codons (RSCU). Histograms reflect overall GC content in all orthologues (gray bar) and by codon position (black bars). GC3 is the most variable between species.

GC Bias in Coding Sequences Affect Codon Usage
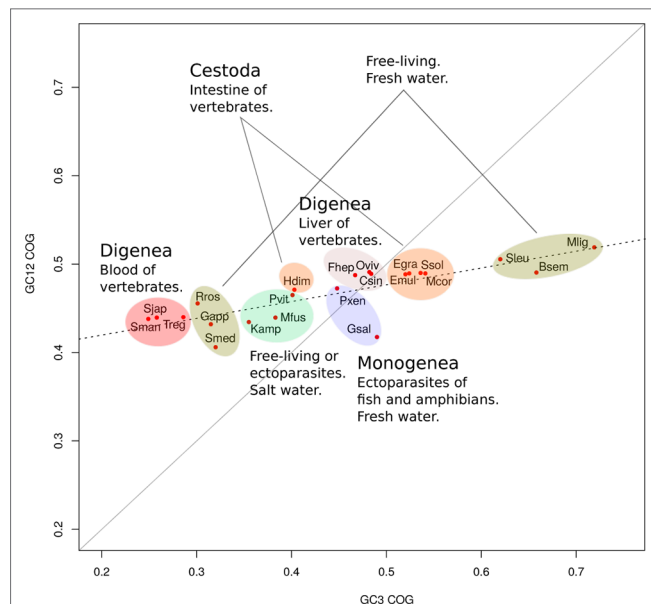
Codon usage bias is a general feature of genomes that has been widely associated with GC content (mutational bias) and natural selection (Bernardi and Bernardi, 1985; Sharp et al., 2010; Plotkin and Kudla, 2011). In this context, we decided to study its extent and its relationship with the genomic GC frequency discussed in the previous section. To this end, heatmaps were plotted to visualize the correlations between GC3 and codon usage (Palidwor et al., 2010). While codon usage bias is observed for all species, the more compositionally skewed organisms (the three plots on the right) show more intense correlations, indicating that the phenomena are stronger in these organisms as might be expected. Also, in most cases, the correlation values were positive for GC-ended codons and negative for the AT-ended ones (**Figure 3**). To further characterize this relationship, the distribution of the frequency of synonymous codons was analyzed for all organisms. A dramatic split of GC- vs AT-ended codons is observed in the species with the more biased GC genomes as the AT-rich model trematode *S. mansoni*. Notably, the split is seen in

opposite directions in the free-living flatworms *S. mediterranea* and *M. lignano* that are at the extremes of the GC distribution (**Figure 4**). A less marked but significant difference is seen within the cestodes consistent with a more balanced GC content, a feature confirmed in the analysis of the 22 species across flatworm diversity (**Supplementary Figure 2**). These observations suggest that genome-wide mutational bias is a major contributor to the observed codon frequency profiles for each organism.

The GC3 vs ENC charts for the analyzed species (**Supplementary Figure 3**) show a combined contribution of selection and mutation for most of the species supporting the trends observed previously, while heavily biased genomes of Schistosomes fall on the curve, suggesting a strong effect of mutational bias.

Differential Codon Usage Is Associated With Expression Levels

While mutation bias influences codon usage in a genome-wide fashion, selection may also act on coding sequences to select for specific codons. This theory predicts that more frequent codons are actually more efficient and/or more accurate during translation of

FIGURE 2 | GC variation and lifestyles correlations. Mean of GC12 versus GC3 for orthologue groups. GC3 is a better discriminator of species than GC12. GC content was linked to the ecological niche of the adult form.

the mRNA (Sharp et al., 2010; Plotkin and Kudla, 2011). To test if this phenomenon is observed in flatworms, steady-state mRNA levels for the available species were collected to differentiate high- and low-expression genes. As shown in **Figure 5**, where the two main components of a PCA of codon usage for *S. mansoni* are plotted, high- and low-expression genes do present a distinct usage profile. Interestingly, when comparing the 10% higher- and lower-expressed
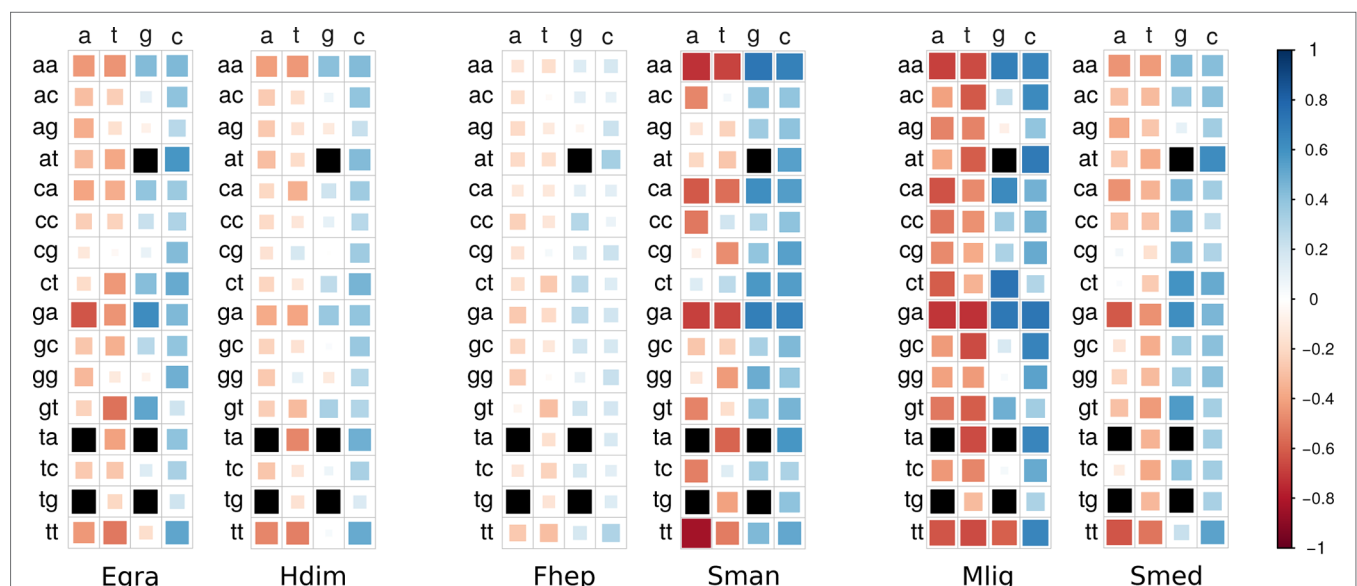
genes in the adult stage, a preference for using GC-rich codons is observed narrowing the distribution in the highly expressed genes and extending it in the lowly expressed (**Supplementary Figure 4**). Even though these results may be explained by biased repair mechanisms acting on highly transcribed sequences, this result is also compatible with translational selection acting on these genes to drive the observed bias.

Amino Acid Usage Is Also Biased in Diverse Flatworm Lineages
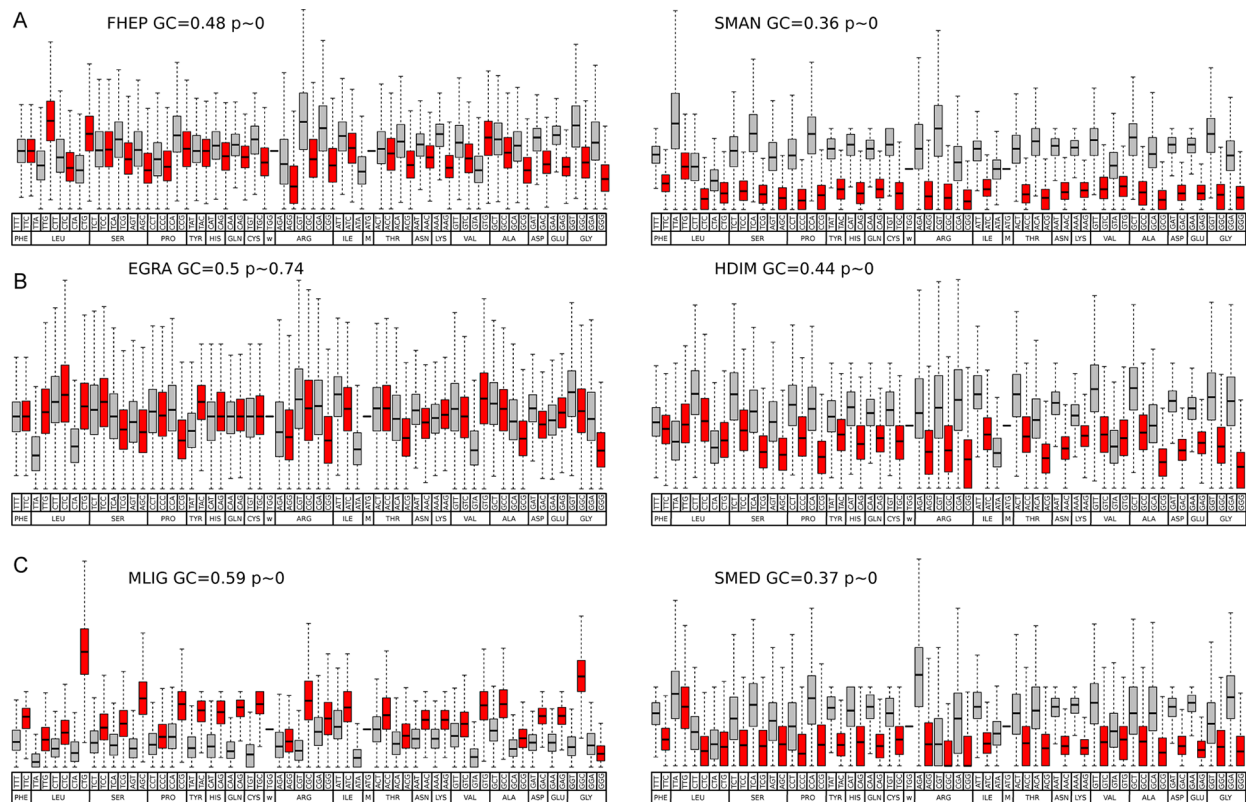
The strong bias observed in codon usage is expected to be associated with the 3rd codon position allowing synonymous changes. However, variations might also exist at the amino acid level (Li et al., 2015). To investigate this, we analyzed the amino acid usage within the set of 700 orthologue genes in pairs of species. Subtle but significant differences in the amino acid frequencies can be detected in cestodes and trematodes mainly involving the amino acids encoded by AT-rich [Ile (AUR), Asn (AAY), Lys (AAR)] or GC-rich [Arg (CGN), Ala (GCN)] codons (**Figure 6** and **Supplementary Table 1**). The variations are more pronounced in the comparison of the free-living species, and in all the cases, the variation follows the GC trend of the species. Since these results are based on a set of orthologue genes, the variations in amino acid frequencies indicate that not only synonymous changes account for the variability observed but also non-synonymous substitutions are taking place.

Synonymous and Non-Synonymous Substitutions in Conserved Orthologous Genes
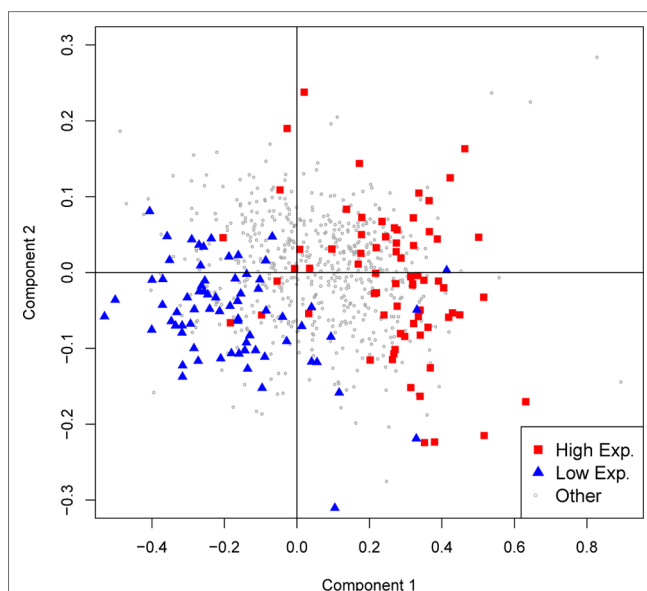
We decided to investigate if particular directional changes could be detected when analyzing orthologous positions in the three

FIGURE 3 | Heatmaps of correlations between codon usage and GC3. Correlation values are coded according to the color scale depicted in the side bar (blue = positive, red = negative). Color intensity and the size of the rectangle are proportional to the correlation coefficients. Black squares: ATG (Met), TGG (Trp), and STOP codons. Organism name abbreviations are as in **Table 1**.

FIGURE 4 | Codon usage in representative species of flatworms. Boxplot of codon usage in six representative species of the diverse lineages and GC content across flatworms. **(A)** comparison of the trematodes *F. hepatica* and *S. mansoni*; **(B)** comparison of the cestodes *E. granulosus* and *H. diminuta*; **(C)** comparison of the free-living *M. lignano* and *S. mediterranea*. GC- and AT-ended codons are grey and red coded for ease of visualization. Mean GC values and p values are indicated.
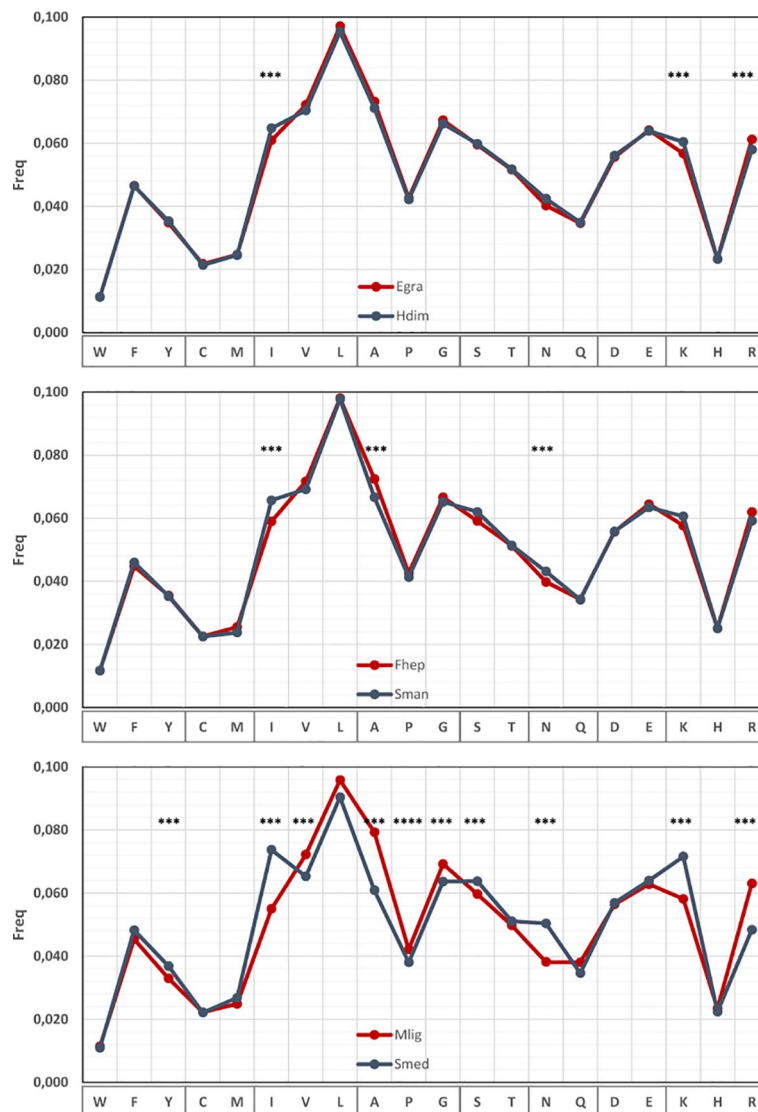
FIGURE 5 | GC variation and expression level in *S. mansoni*. The first two axes of a principal component analysis on codon frequencies are plotted. Genes with expression levels above the 90th percentile or below the 10th percentile are colored red and blue, respectively.

paired species. For this, we selected the ungapped regions of each pairwise alignment of orthologues and generated substitution matrixes based on the aligned orthologous positions. Amino acid conservation was generally high, with tryptophan (Trp) and glycine (Gly) as the more conserved residues in all the species, confirming Ile, Ser, Ala, Asn, and Arg as the more variable (**Supplementary Table 2**). Expectedly, the most frequent changes involved amino acids with similar properties (**Figure 7A**), particularly those involving aliphatic and hydrophilic residues. However, several reciprocal changes showed significant differences in the counts (**Figure 7B**). Similar effects can be observed in the other pairwise comparisons (**Supplementary Figure 5**).

To gain further insights into these phenomena, we evaluated the substitutions at the codon level generating substitutions matrices for the three pair of species selected (**Supplementary Table 3**).

The conservation at codon level as expected was much lower with a strong component of synonymous changes (**Supplementary Table 3**). The lower frequency of GC-rich codons in *S. mansoni* (depicted in **Figure 4**) is explained by a marked increase of synonymous substitutions toward AT-rich codons (**Figure 8**). Similarly, in *M. lignano*, synonymous substitutions toward GC-rich codons are associated with reduced AT codon counts (**Supplementary Figure 6**).

FIGURE 6 | Amino acid usage differences within flatworms. Amino acid usage in a set or 700 orthologous genes compared in pairs of species with different global GC. The amino acids showing significant differences are indicated.
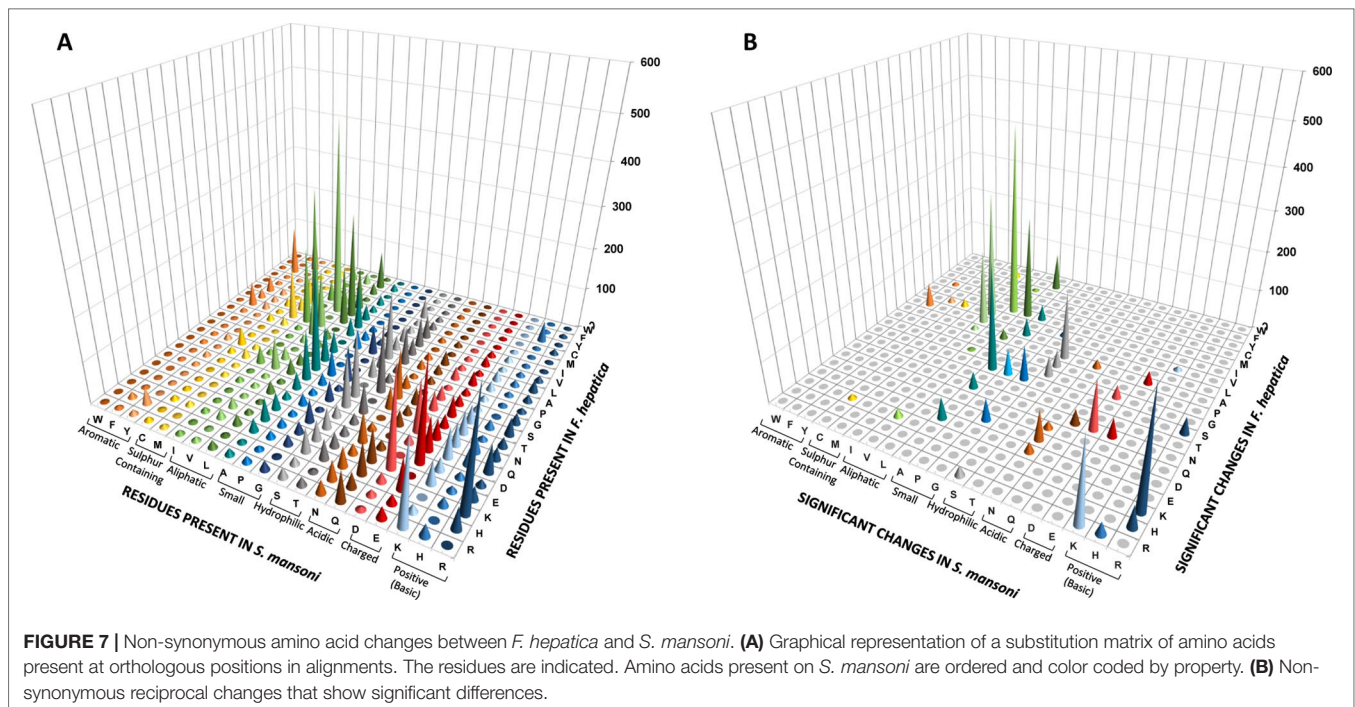
The analysis of the non-synonymous changes at the codon level showed an increased complexity (**Supplementary Table 3**). One striking feature is that amino acid changes involving two substitutions are more common than those explained by simple substitutions. A detailed example is presented in **Figure 9**.

Ala is a relatively GC-rich codon (GCN) that is frequently substituted by the more GC neutral Ser (TCN + AGY) and vice versa (**Figures 7** and **9**). When this substitution takes place, it is expected that the GCN codon would change for the corresponding TCN variant, i.e., that GCA would turn into TTA and GCG into TCG. In 49 positions in the alignment, a GCG coding Ala is present in *F. hepatica*, while Ser codons are present in *S. mansoni* (second row). The simple transversion GCG to TCG is underrepresented with only six occurrences,

while the changes toward TCA and TCT are more abundant (23 and 15 occurrences, respectively). Similarly, the GCC to TCC transversion (third row) represents only a 15% of the Ala (GCC) changes to Ser, while the more AT-rich variants (TCA and TCT) represent more than 56% of the substitutions. Notably, when the AT-ending Ala codons (GCA and GCT, 1st and 4th rows) are substituted, the more common codon is the one expected by a single 3rd position change.

The reciprocal Ser (in *F. hepatica*) to Ala (in *S. mansoni*) changes are less common (169 times vs 263 Ala to Ser), but again, in 77% of the cases, they are enforced to 3rd position T or A irrespective of the original Ser codon. Similar effects can be seen when analyzing other amino acid changes (**Supplementary Table 3**) and particularly in those regarded as significant (from **Figure 7**) as Ile to Val or Lys to Arg (**Supplementary Figure 7**).

FIGURE 7 | Non-synonymous amino acid changes between *F. hepatica* and *S. mansoni*. **(A)** Graphical representation of a substitution matrix of amino acids present at orthologous positions in alignments. The residues are indicated. Amino acids present on *S. mansoni* are ordered and color coded by property. **(B)** Non-synonymous reciprocal changes that show significant differences.

Taken together, these results are strongly suggestive of a combined effect of mutation and selection in order to maintain both the compositional GC skew of the species and the property of the coded amino acid. In other words, whenever an amino acid change occurs through a simple substitution, this is then rapidly switched to those that follow the GC of the species.

DISCUSSION
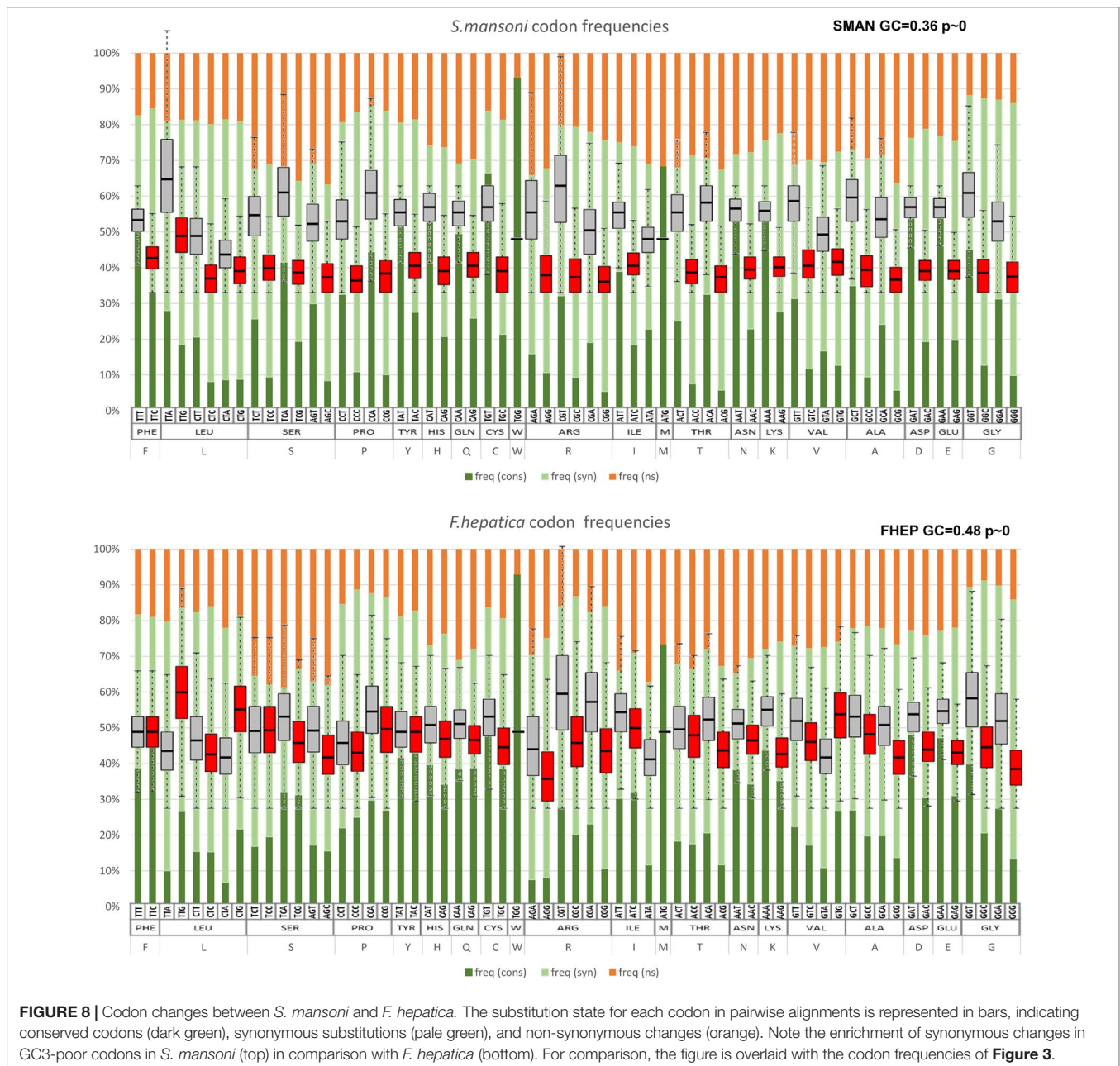
Platyhelminthes classes show a wide range of GC composition, even within groups. Our results show that GC3 content explains most of the observed variability in the codon usage as reflected by the variation in the RSCU values. Based on GC3 variability, we found different clusters within the free-living species, trematodes, and cestodes. This can be clearly seen when the species tree is compared with the tree representing codon usage similarity. Indeed, Platyhelminthes show great differences between both trees, while this phenomenon is not seen in other models as different as bacteria and hexapoda (Behura and Severson, 2012; Dilucca et al., 2018). A similar study in nematodes show a comparably wide distribution of GC values, although the variations are more consistent with the phylogeny (Cutter et al., 2006). These results suggest more recent and strong compositional shifts for these groups of organisms. Further work is needed to explain this particular phenomenon in flatworms.

Codon usage bias is a generalized feature of the genomes of many organisms that is deeply influenced by evolutionary phenomena and results basically from the balance between mutational bias and natural selection (see Plotkin and Kudla, 2011, for a review). To assert the relative influence of these two factors, a plot of GC1–2 vs GC3 for all the species taken together was generated. This "pseudo" neutrality plot shows a slight slope, indicating that GC3 shows a different behavior when compared with GC1–2, a result that is generally taken as evidence that selection is acting to shape codon usage (Sueoka, 1988). Furthermore, this plot shows clearly distinctive clusters within trematodes, cestodes, and free-living species based on GC content. Notably, these differences seem to blurrily reflect the diversity of lifestyles and niches of the diverse flatworms. The observed differences question the use of single species as a model for each class; a clear demonstration of this is the differences between the model *S. mansoni* and other trematodes observed in the boxplot of **Figure 4**. Codon usage bias in flatworm mitochondrial genomes has also been reported (Le et al., 2004; Mazumder et al., 2018). Even though considering genic and genomic large differences, they may follow evolutionary pressures independent from the nuclear genomes.

Interestingly, a similar study across nematodes found robust evidence for selection on codon usage bias in free-living species, a feature found marginally in parasitic ones, and particularly in the most compositionally biased (Cutter et al., 2006). The association of selective bias in free-living or parasitic species is not clear-cut in the case of flatworms, which might be reflecting diverse evolutionary strategies.

In agreement with the hypothesis of translational selection driving synonymous codon usage bias, we observe a clear association of gene expression levels with codon usage where highly expressed genes are rich in GC-rich codons, while the opposite is observed for low-expression genes. Similar results have been previously reported for cestodes, among others (Chen et al., 2013; Yang et al., 2014; Yang et al., 2015; Huang et al., 2017; Maldonado et al., 2018). Even in highly AT-biased genomes—as observed for the schistosomes—the GC content

FIGURE 8 | Codon changes between *S. mansoni* and *F. hepatica*. The substitution state for each codon in pairwise alignments is represented in bars, indicating conserved codons (dark green), synonymous substitutions (pale green), and non-synonymous changes (orange). Note the enrichment of synonymous changes in GC3-poor codons in *S. mansoni* (top) in comparison with *F. hepatica* (bottom). For comparison, the figure is overlaid with the codon frequencies of **Figure 3**.

of highly expressed genes is relatively high when compared with that of the general trend. It is worth to mention that the bias in repair mechanisms of actively transcribed DNA has also being proposed to explain this observation.

The observed differences in CG and codon usage among these organisms are also reflected in the amino acid composition. Recently, Li et al. (2015) show the strong relationship of synonymous codon usage and differential amino acid usage, using a strategy based on classifying amino acids in three groups (high, medium, and low GC content) according to the GC composition of their corresponding codons. Our results on amino acid frequencies in the different flatworm species are consistent with these observations.

Furthermore, when orthologues positions are considered, we mainly observed amino acid substitutions conservative of the physicochemical properties as would be expected. However, these changes frequently involve codons of completely different GC content that follow the differences observed in general GC content of the genomes, i.e., Ile (AUH) vs Leu (CCN, UCR). In this way, AT-rich genomes accumulate changes to amino acids in the low GC group, while the opposite is observed in GC-rich genomes.

Remarkably, when the frequency of a certain amino acid substitution is not reciprocal between two given organisms, the amino acids involved belong to the different groups defined by Li et al. (2015). An interesting case is observed for the Lys to

Fhep \ Sman		A				S					
		GCA	GCG	GCC	GCT	TCA	TCG	TCC	TCT	AGC	AGT
A	GCA	204	45	96	261	21	9	11	18	1	9
	GCG	124	32	58	167	23	6	3	15	1	1
	GCC	205	31	86	318	25	10	11	17	3	8
	GCT	269	62	101	396	21	8	8	25	3	6
S	TCA	9	3	2	10	140	31	41	98	4	9
	TCG	12	2	4	17	91	34	24	85	3	18
	TCC	14	4	5	19	132	33	50	134	7	7
	TCT	15	4	6	17	117	33	38	122	3	9
	AGC	0	3	1	7	19	1	2	10	81	155
	AGT	4	2	2	7	13	3	4	18	74	215

FIGURE 9 | Non-synonymous changes between *S. mansoni* and *F. hepatica*. Detail of the substitution matrix of **Supplementary Table 3** of the changes involving Ala and Ser (*S. mansoni* codons in columns, *F. hepatica* in rows). Note the lower-than-expected counts of changes toward GC3-rich codons and the enrichment in synonymous substitutions that imply two substitutions.

Arg substitution. Even though these amino acids have similar physicochemical properties, they belong to opposite groups, being Lys coded by the most AT-rich group of codons, while Arg is on the highest GC content side.

A detailed analysis of the non-synonymous changes showed a higher-than-expected frequency of codon changes involving two nucleotides. This is paralleled by a marked reduction in the counts of the expected codon substitutions involving a single change. A plausible explanation for this phenomenon is offered by a two-hit mechanism, providing a clear example of the combined effect of mutation and selection. The two-hit hypothesis proposed implies that when a mutation changes the coded amino acid, this non-synonymous substitution is rapidly adapted to the general GC content of the genome by a second synonymous change.

CONCLUSIONS

GC bias has a great influence on synonymous codon and amino acid usage across Platyhelminthes, a feature not shared by all metazoans. Both free-living and parasitic species show the phenomena, and no clear correlation with lifestyles or evolutionary closeness is evident so far. The changes introduced by GC bias impact not only in synonymous codon usage but also in amino acid frequencies. The evidence so far suggests that both mutation and selection are acting to shape the coding strategies of the diverse flatworms.

DATA AVAILABILITY

All datasets generated for this study are included in the manuscript/supplementary files.

AUTHOR CONTRIBUTIONS

GL and SF performed the bioinformatics analysis and contributed in writing the manuscript. GR performed bioinformatics analysis. PS and JT participated in the design of the study and the interpretation of data, drafting the manuscript, and critical revision of its content. All authors read and approved the final manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00771/full#supplementary-material.

REFERENCES

Alvarez, F., Garat, B., Musto, H., Picon, M., and Ehrlich, R. (1993). Tendencies in echinococcus sp. codon usage. *Mem. Inst. Oswaldo Cruz* 88, 345–346. doi: 10.1590/S0074-02761993000200029

Behura, S. K., and Severson, D. W. (2012). Comparative analysis of codon usage bias and codon context patterns between dipteran and hymenopteran sequenced genomes. *PLoS One* 7 (8), e43111.

Behura, S. K., and Severson, D. W. (2013). Codon usage bias: causative factors, quantification methods and genome-wide patterns: with emphasis on insect genomes. *Biol. Rev.* 88 (1), 49–61. doi: 10.1111/j.1469-185X.2012.00242.x

Bernardi, G., and Bernardi, G. (1985). Codon usage and genome composition. *J. Mol. Evol.* 22, 363–365. doi: 10.1007/BF02115693

Caira, J. N., and Littlewood, D. T. J. (2013). "Worms, platyhelminthes," in *Encyclopedia of Biodiversity: Second Edition* (USA, MA: Academic Press) 437–469. doi: 10.1016/B978-0-12-384719-5.00166-0

Charif, D., and Lobry, J. R. (2007). "SeqinR 1.0-2: a contributed package to the r project for statistical computing devoted to biological sequences retrieval and analysis," in *Structural approaches to sequence evolution: Molecules, networks, populations*. Eds. U. Bastolla, M. Porto, H. E. Roman, and M. Vendruscolo, Biological and Medical Physics, Biomedical Engineering (New York: Springer Verlag), 207–232. doi: 10.1007/978-3-540-35306-5_10

Chen, L., Liu, T., Yang, D., Nong, X., Xie, Y., Fu, Y., et al. (2013). Analysis of codon usage patterns in Taenia pisiformis through annotated transcriptome data. *Biochem. Biophys. Res. Commun.* 430, 1344–1348. doi: 10.1016/j.bbrc.2012.12.078

Coghlan, A., Tyagi, R., Cotton, J. A., Holroyd, N., Rosa, B. A., Tsai, I. J., et al. (2019). Comparative genomics of the major parasitic worms. *Nat. Genet.* 51, 163–174. doi: 10.1038/s41588-018-0262-1

Collins, J. J., 3rd (2017). Platyhelminthes. *Curr. Biol.* 27 (7), R252–R256. doi: 10.1016/j.cub.2017.02.016

Cutter, A. D., Wasmuth, J. D., and Blaxter, M. L. (2006). The evolution of biased codon and amino acid usage in nematode genomes. *Mol. Biol. Evol.* 23, 2303–2315. doi: 10.1093/molbev/msl097

Dilucca, M., Cimini, G., and Giansanti, A. (2018). Essentiality, conservation, evolutionary pressure and codon bias in bacterial genomes. *Gene.* 663, 178–188. doi: 10.1016/j.gene.2018.04.017

Egger, B., Lapraz, F., Tomiczek, B., Müller, S., Dessimoz, C., Girstmair, J., et al. (2015). A transcriptomic-phylogenomic analysis of the evolutionary relationships of flatworms. *Curr. Biol.* 25, 1347–1353. doi: 10.1016/j.cub.2015.03.034

Ellis, J., Morrison, D. A., and Kalinna, B. (1995). Comparison of the patterns of codon usage and bias between Brugia, Echinococcus, Onchocerca and Schistosoma species. *Parasitol. Res.* 81, 388–393. doi: 10.1007/BF00931499

Ellis, J. T., and Morrison, D. A. (1995). Schistosoma mansoni: patterns of codon usage and bias. *Parasitology* 110, 53–60. doi: 10.1017/S003118200008104X

Fernandez, V., Zavala, A., and Musto, H. (2001). Evidence for translational selection in codon usage in Echinococcus spp. *Parasitology* 123, 203–209. doi: 10.1017/S0031182001008150

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010

Howe, K. L., Bolt, B. J., Shafie, M., Kersey, P., and Berriman, M. (2017). WormBase ParaSite – a comprehensive resource for helminth genomics. *Mol. Biochem. Parasitol.* 215, 2–10. doi: 10.1016/j.molbiopara.2016.11.005

Huang, X., Xu, J., Chen, L., Wang, Y., Gu, X., Peng, X., et al. (2017). Analysis of transcriptome data reveals multifactor constraint on codon usage in Taenia multiceps. *BMC Genomics* 18, 1–13. doi: 10.1186/s12864-017-3704-8

Kalinna, B. H., and McManus, D. P. (1994). Codon usage in echinococcus. *Exp. Parasitol.* 79, 72–76. doi: 10.1006/expr.1994.1063

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Lamolle, G., Protasio, A. V., Iriarte, A., Jara, E., Simón, D., and Musto, H. (2016). An isochore-like structure in the genome of the flatworm schistosoma mansoni. *Genome Biol. Evol.* 8, 2312–2318. doi: 10.1093/gbe/evw170

Larsson, K., and Jondelius, U. (2008). Phylogeny of Catenulida and support for Platyhelminthes. *Org. Divers. Evol.* 8, 378–387. doi: 10.1016/j.ode.2008.09.002

Laumer, C. E., and Giribet, G. (2014). Inclusive taxon sampling suggests a single, stepwise origin of ectolecithality in Platyhelminthes. *Biol. J. Linn. Soc.* 111, 570–588. doi: 10.1111/bij.12236

Laumer, C. E., Hejnol, A., and Giribet, G. (2015). Nuclear genomic signals of the "microturbellarian" roots of platyhelminth evolutionary innovation. *Elife* 4, 1–31. doi: 10.7554/eLife.05503

Le, T. H., McManus, D. P., and Blair, D. (2004). Codon usage and bias in mitochondrial genomes of parasitic platyhelminthes. *Korean J. Parasitol.* 42 (4), 159–167. doi: 10.3347/kjp.2004.42.4.159

Li, J., Zhou, J., Wu, Y., Yang, S., and Tian, D. (2015). GC-content of synonymous codons profoundly influences amino acid usage. *G3 Genes Genomes Genet.* 5, 2027–2036. doi: 10.1534/g3.115.019877

Maldonado, L. L., Stegmayer, G., Milone, D. H., Oliveira, G., Rosenzvit, M., and Kamenetzky, L. (2018). Whole genome analysis of codon usage in Echinococcus. *Mol. Biochem. Parasitol.* 225:54-66. doi: 10.1016/j.molbiopara.2018.08.001

Mazumder, G. A., Uddin, A., and Chakraborty, S. (2017a). Comparative analysis of codon usage pattern and its influencing factors in Schistosoma japonicum and Ascaris suum. *Acta Parasitol.* 62, 748–761. doi: 10.1515/ap-2017-0090

Mazumder, G. A., Uddin, A., and Chakraborty, S. (2017b). Expression levels and codon usage patterns in nuclear genes of the filarial nematode Wuchereria bancrofti and the blood fluke Schistosoma haematobium. *J. Helminthol.* 91, 72–79. doi: 10.1017/S0022149X16000092

Mazumder, G. A., Uddin, A., and Chakraborty, S. (2018). Codon usage pattern of complex III gene of respiratory chain among platyhelminths. *Infect. Genet. Evol.* 57, 128–137. doi: 10.1016/j.meegid.2017.10.018

Meadows, H. M., and Simpson, A. J. G. (1989). Codon usage in Schistosoma. *Mol. Biochem. Parasitol.* 36, 291–293. doi: 10.1016/0166-6851(89)90178-3

Milho, J. L., and Tracy, J. W. (1995). Updated codon usage in schistosoma. *Exp. Parasitol.* 80, 353–356. doi: 10.1006/expr.1995.1046

Mitreva, M., Wendl, M. C., Martin, J., Wylie, T., Yin, Y., Larson, A., et al. (2006). Codon usage patterns in Nematoda: analysis based on over 25 million codons in thirty-two species. *Genome Biol.* 7 (8), R75. doi: 10.1186/gb-2006-7-8-r75

Musto, H., Romero, H., and Rodríguez-Maseda, H. (1998). Heterogeneity in codon usage in the flatworm Schistosoma mansoni. *J. Mol. Evol.* 46, 159–167. doi: 10.1007/PL00006291

Palidwor, G. A., Perkins, T. J., and Xia, X. (2010). A general model of codon bias due to GC mutational bias. *PLoS One* 5 (10), e13431. doi: 10.1371/journal.pone.0013431

Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289–290. doi: 10.1093/bioinformatics/btg412

Plotkin, J. B., and Kudla, G. (2011). Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.* 12, 32–42. doi: 10.1038/nrg2899

Protasio, A. V., Tsai, I. J., Babbage, A., Nichol, S., Hunt, M., Aslett, M. A., et al. (2012). A systematically improved high quality genome and transcriptome of the human blood fluke schistosoma mansoni. *PLoS Negl. Trop. Dis.* 6 (1), e1455. doi: 10.1371/journal.pntd.0001455

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

Sharp, P. M., Emery, L. R., and Zeng, K. (2010). Forces that influence the evolution of codon bias. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 365, 1203–1212. doi: 10.1098/rstb.2009.0305

Sueoka, N. (1988). Directional mutation pressure and neutral molecular evolution. *Proc. Natl. Acad. Sci.* 85, 2653–2657. doi: 10.1073/pnas.85.8.2653

Wei, T., and Simko, V. (2017). R package "corrplot": visualization of a correlation matrix (version 0.84). Available at: https://github.com/taiyun/corrplot.

Wright, F. (1990). The effective number of codons used in a gene. *Gene.* 87, 23–29. doi: 10.1016/0378-1119(90)90491-9

Yang, X., Luo, X., and Cai, X. (2014). Analysis of codon usage pattern in Taenia saginata based on a transcriptome dataset. *Parasit. Vectors* 7, 1–11. doi: 10.1186/s13071-014-0527-1

Yang, X., Ma, X., Luo, X., Ling, H., Zhang, X., and Cai, X. (2015). Codon usage bias and determining forces in Taenia solium genome. *Korean J. Parasitol.* 53, 689–697. doi: 10.3347/kjp.2015.53.6.689

frontiers
in Genetics

Males, the Wrongly Neglected Partners of the Biologically Unprecedented Male–Female Interaction of Schistosomes

Zhigang Lu[1,2], Sebastian Spänig[3], Oliver Weth[2] and Christoph G. Grevelding[2]*

[1] Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, United Kingdom, [2] Insitute for Parasitology, BFS, Justus Liebig University Giessen, Giessen, Germany, [3] Department of Mathematics & Computer Science, University of Marburg, Marburg, Germany

Schistosomes are the only platyhelminths that have evolved separate sexes, and they exhibit a unique reproductive biology because the female's sexual maturation depends on a constant pairing contact with the male. In the female, pairing leads to gonad differentiation, which is associated with substantial morphological changes, and controls among others the expression of gonad-associated genes. In the male, no morphological changes have been observed after pairing, although first data indicated an effect of pairing on gene transcription. Comprehensive transcriptomic approaches have revealed an unexpected high number of genes that are differentially transcribed in the male after pairing. Their identities suggest roles for the male that are not restricted to feeding and enhanced muscular power to transport paired female and, as assumed before, to induce its sexual maturation by one "magic" factor. Instead, a more complex picture emerges in which both partners live in a reciprocal sender-recipient relationship that not only affects the gonads of both genders but may also involve tactile stimuli, transforming growth factor β signaling, nutritional parts, and neuronal processes, including neuropeptides and G protein-coupled receptor signaling. This review provides a summary of transcriptomics including an overview of genes expressed in a pairing-dependent manner in schistosome males. This may stimulate further research in understanding the role of the male as the recipient of the female's signals upon pairing, the male's "capacitation," and its subsequent competence as a sender of information. The latter process finally transforms a sexually immature, autonomous female without completely developed gonads into a sexually mature, partially non-autonomous female with fully differentiated gonads and enormous egg production capacity.

Keywords: schistosomes, male–female interaction, transcriptomics, pairing-dependent gene expression, TGFβ signaling, neuropeptide, G protein-coupled receptor

SCHISTOSOMES AND THE MALE–FEMALE INTERACTION

Schistosomes are parasitic platyhelminths causing schistosomiasis (bilharzia), an infectious disease of worldwide importance for humans and animals. The World Health Organization has listed schistosomiasis as one of the neglected tropical diseases. More than 200 million people required preventive treatment in 2016 (World Health Organization, 2019; McManus et al., 2018). This disease is

most prevalent in Africa but occurs also in Asia and South America due to the presence of tropical water snails as intermediate hosts, which prefer warm (sub)tropical habitats. Unexpectedly, there is recent evidence of an autochthonous site also in southern Europe (Corsica), where the snail-host species occurs due to moderate climate conditions (Boissier et al., 2016).

As a waterborne disease, schistosomiasis affects humans and animals exposed to water infested with cercariae, the infectious larval stage originating from snails. Some schistosome species comprise zoonotic potential, which increases the risk of infection (Standley et al., 2012). Together, these facts make disease control difficult and contribute to an additional, socioeconomic problem (Garchitorena et al., 2017).

The pathology of schistosomiasis is triggered by eggs that paired females deposit in the bloodstream of vertebrate hosts. These eggs eventually lodge in organs such as the liver where they cause inflammation and fibrosis (Olveda et al., 2014; McManus et al., 2018). The prerequisite for egg production is the complete development of the female gonads. This, however, is only achieved if a constant pairing contact with a male has been established. To this end, the female resides within a ventral groove formed by the male, the gynecophoral canal. This close liaison can last over years, an exceptional phenomenon in nature (Basch, 1991; Grevelding, 2004). Males play a pivotal role in controlling schistosome reproduction by inducing mitoses and differentiation processes in the reproductive organs (ovary and vitellarium, the latter providing vitelline cells for the production of mature eggs) of the paired female, a process that comes along with a significant increase of its body size (Popiel and Basch, 1984a; Den Hollander and Erasmus, 1985; Kunz, 2001). Pairing even controls the expression of female-specific expressed genes with functions in the vitellarium (LoVerde and Chen, 1991; Grevelding et al., 1997). Although the molecular consequences of pairing on females have been a strong focus of basic research (Hoffmann, 2004; LoVerde et al., 2009; Beckmann et al., 2010), pairing-dependent processes in males are somewhat neglected.

A HISTORICAL SNAPSHOT OF THE MALE SCHISTOSOMES' PERSPECTIVE

With respect to their sexual biology, schistosome males appear "ready to go." Being independent of pairing, they possess fully developed testes and seminal vesicles filled with differentiated sperms as confirmed by morphological analyses (Neves et al., 2005; Beckmann et al., 2010). Sperm was excluded as a factor inducing the sexual maturation of the female, and irradiated or surgically manipulated males lacking testes were still capable of mating and inducing sexual maturation in paired females as well as egg production (Armstrong, 1965; Michaels, 1969). Alternatively, the male was proposed to deliver specific molecules during pairing that supervise body length and the sexual maturation of the female (Armstrong, 1965; Basch and Basch, 1984), which depends on the pairing status; unpaired females are infertile because their reproductive organs have not fully developed (Neves et al., 2005; Beckmann et al., 2010). Furthermore, female sexual maturation was hypothesized to be a consequence of local activities of molecules (Michaels, 1969; Popiel and Basch, 1984b). In addition, a tactile

impulse was proposed (Basch and Basch, 1984). Finally, a male-secreted hormonal factor(s) was suggested to be transferred to the female (Ruppel and Cioli, 1977; Shaw et al., 1977; Atkinson and Atkinson, 1980; Basch and Nicolas, 1989). However, none of these hypotheses resulted in the identification of the "magic male factor".

From the metabolic perspective, glucose and cholesterol were demonstrated to be delivered by the male during pairing, and it was hypothesized that nourishment contributes to female sexual maturation (Conrford and Huot, 1981; Cornford and Fitzpatrick, 1985; Haseeb et al., 1985; Silveira et al., 1986). The most persuasive evidence for an important player in the game resulted from studies about the gynecophoral canal protein (GCP). First detected in adult *Schistosoma mansoni*, *Sm*GCP was identified as a glycoprotein putatively transferred from the male to the female (Gupta and Basch, 1987) and later, by immunolocalization, to be widely distributed on the surface of a paired female (Bostic and Strand, 1996). Structurally, *Sm*GCP lacks a transmembrane domain but reveals short, conserved repeat regions with sequence similarity to fasciclin I, a neuronal cell-adhesion protein. In males, *Sm*GCP expression appeared to be limited to the gynecophoral canal region, the mating partners' interface. Furthermore, *Sm*GCP seemed to be down-regulated in unpaired males. These findings suggested that *Sm*GCP is diffusible and delivered by the male during a pairing contact (Bostic and Strand, 1996). Indeed, results of a subsequent study in *Schistosoma japonicum* indicated the importance of *Sj*GCP for pairing. RNA interference experiments against *Sj*GCP resulted in reduced pairing stability *in vitro* and *in vivo* (Cheng et al., 2009). Finally, evidence was found for the regulation of *Sm*GCP *via* transforming growth factor β (TGFβ)-dependent signaling in *S. mansoni* (Osman et al., 2006). Although the biochemical activity of GCP has not yet been clearly addressed, there is accumulating evidence for its participation in male–female interaction.

In earlier studies, the DNA synthesis marker [³H]thymidine was used in incorporation assays with females *in vitro* to determine the mitosis rates dependent on the pairing. Comparing females paired in the presence of thymidine to either pairing-experienced males (bM, bisex males) or pairing-inexperienced males (sM, single-sex males) demonstrated that maturity is decisive. To induce mitogenic activity in females, sM required a significantly longer mating period (≥24 h) than bM (Den Hollander and Erasmus, 1985), which stimulated mitogenic activity in females within the first 24 h of pairing. This early study already pointed toward bidirectional communication between the partners during the initial phase of pairing. Furthermore, this result suggests that males have to pass through a process of capacitation before they acquire competence to supervise female sexual maturation — part of which is the induction of mitoses (Knobloch et al., 2002).

TRANSCRIPTOMIC PERFORMANCE OF MALE SCHISTOSOMES

During the last 15 years, *omic* studies have allowed unprecedented insights in the life processes of a great variety of organisms (Weissenbach, 2016), including schistosomes (Verjovski-Almeida et al., 2003; Hu et al., 2003; Berriman et al., 2009; Schistosoma japonicum Genome Sequencing and Functional Analysis

Consortium, 2009; Protasio et al., 2012; Young et al., 2012; Anderson et al., 2015; Smit et al., 2015; Cai et al., 2016; Sotillo et al., 2017; Wang et al., 2017; Giera et al., 2018). Whereas the majority of these studies applied RNA-seq techniques, microarray analyses and serial analysis of gene expression (SAGE/SuperSAGE) were alternatively used. Among others, these techniques were also applied to compare bM and sM. One SAGE-based approach found differential regulation for transcripts contributing to developmental processes, metabolism, and the redox system (Williams et al., 2007). Even before the genome project was finished, an early microarray analysis found 30 genes to be exclusively transcribed in bM and 66 in sM (Fitzpatrick and Hoffmann, 2006). The identities of these differentially expressed genes indicated their involvement in RNA metabolic processes, which was independently supported in another microarray study (Waisberg et al., 2007). In another approach combining SuperSAGE (a second-generation SAGE technique allowing the identification of longer RNA sequence tags) and microarray analyses, corresponding data sets were produced to get a comprehensive overview of genes differentially transcribed between bM and sM. Among 6326 sense transcripts detected by both analyses, 29 were found to be significantly differentially transcribed (Leutner et al., 2013). Besides differences in the transcript levels of genes involved in metabolic processes, evidence was obtained for additional differences in neuronal processes and TGFβ signaling. In this context, a *S. mansoni* ortholog of follistatin (*Sm*Fst; Smp_123300) was found to be differentially transcribed with an interesting bias toward sM. The latter was independently confirmed by a subsequent RNA-seq approach including paired and unpaired adults and their gonads (Lu et al., 2016; Lu et al., 2017) as well as by independent quantitative polymerase chain reaction (qPCR) analyses (Leutner et al., 2013; Haeberlein et al., 2019). Based on the corresponding results from microarray, SuperSAGE, RNA-seq, and qPCRs, all exhibiting higher transcript occurrence in sM, *Sm*Fst is probably the most intensively studied gene with respect to expression profiling. Thus, it can be used in future studies as a marker for differential transcription in bM versus sM (Haeberlein et al., 2019). Follistatins are known antagonists in TGFβ signaling pathways and block ligands of the TGFβ family such as TGFβ, activin, and bone morphogenetic protein (BMP) (Massagué and Chen, 2000; Moustakas and Heldin, 2009; Heldin and Moustakas, 2016). The first characterization of *Sm*Fst showed testes localization (by *in situ* hybridization). By yeast two-hybrid analyses, an interaction potential with *S. mansoni* orthologs of the TGFβ ligands *Sm*InAct and *Sm*BMP was found. The agonists colocalized with *Sm*Fst in the testes (Leutner et al., 2013). These results suggest that TGFβ signaling also plays a role in male-female interaction and is part of the bidirectional communication between both genders. As such, *Sm*Fst could represent one of several competence factors of males expressed in response to pairing. Indeed, a recent *in vitro* study with paired, separated, and re-paired males demonstrated an immediate influence of pairing on the on/off transcriptional status of *Sm*Fst (Haeberlein et al., 2019). This finding adds to previous hypotheses that TGFβ signaling is involved in pairing-dependent reproductive processes in schistosomes (LoVerde et al., 2007; Buro et al., 2013). One role of *Sm*Fst in sM might be the prevention of the activation of one or more of its TGFβ pathways before pairing. Figuratively seen, *Sm*Fst in sM appears like a systemic handbrake of a specific, male competence-related

biological driving route needed after pairing to reach maturation – a hypothesis that awaits corroboration.

Today, RNA-seq represents the state-of-the-art technology for transcriptome analysis providing both a wide analytical range and the quantification of study samples. Theoretically, RNA-seq can cover all transcripts of a biological sample, an advantage over microarrays or SAGE/SuperSAGE (Marioni et al., 2008). Recently, RNA-seq was applied for comparative transcript profiling in paired and unpaired adult *S. mansoni* and their gonads. Of more than 7,000 transcripts detected in the gonads, 243 (testes) and 3,600 (ovaries) were transcribed in a pairing-dependent manner. In addition to genes preferentially or specifically transcribed in adults and gonads of both genders, evidence was obtained for pairing-dependent processes in the gonads affecting genes with, for example, stem cell-associated functions. This was particularly expected for females due to the pairing-induced differentiation processes in the vitellarium and the ovary (Erasmus, 1973; Shaw, 1987; Kunz, 2001; Neves et al., 2005; Beckmann et al., 2010). Remarkably, from their annotation, many differentially transcribed genes appeared to be involved in neuronal processes. This perception substantiated one of the results of the combinatory SuperSAGE/microarray approach comparing sM and bM transcript profiles (Leutner et al., 2013).
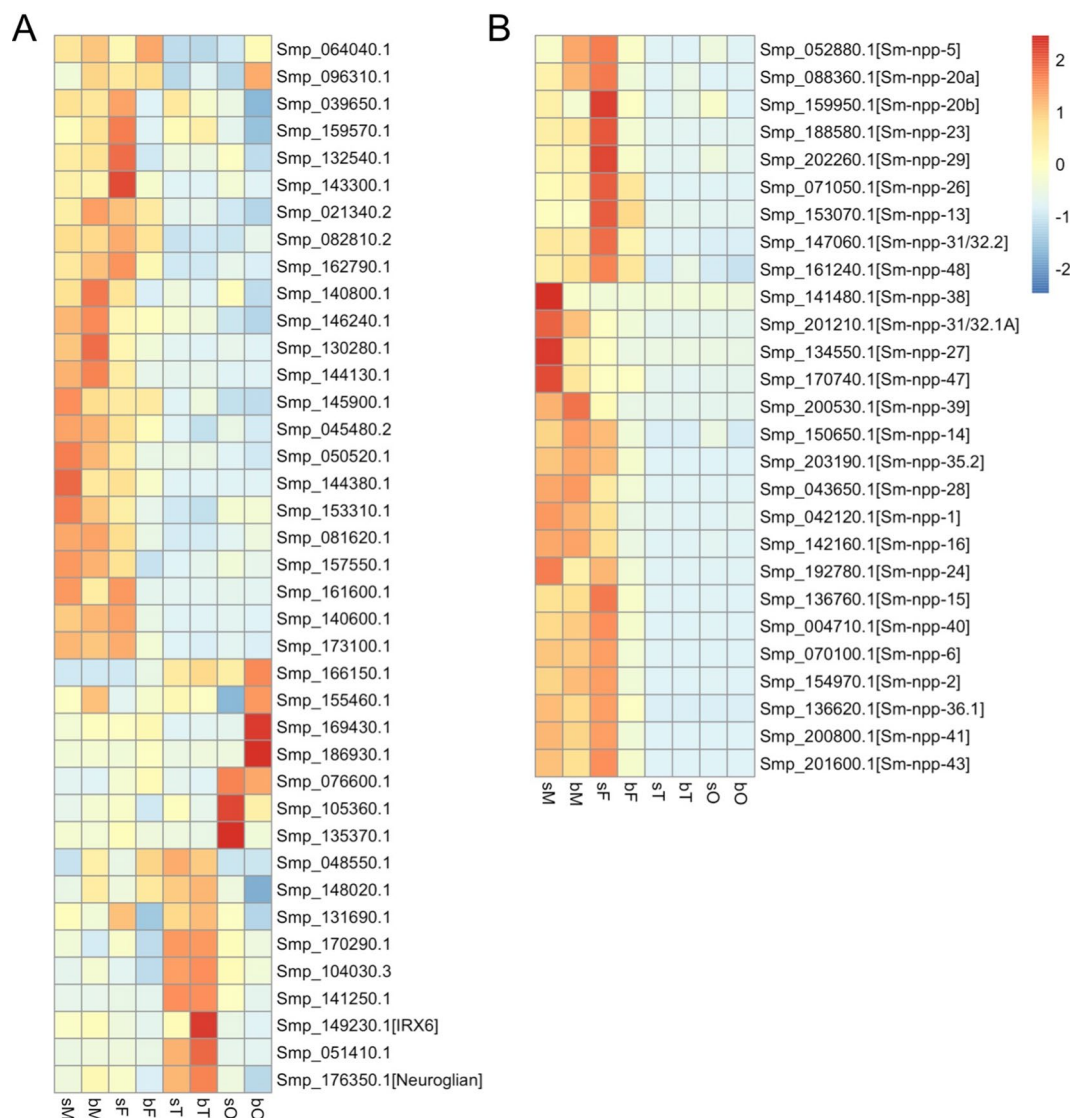
"NERVOUS" MALE SCHISTOSOME

One objective of human neuroscience is to understand how neuronal circuits direct behavior, how humans perceive the world, how they learn from experience, how memory works, how movements are directed, and how communication is realized (National Research Council (US) Committee on Research Opportunities in Biology, 1989). The basis for integrating all these interactions and requirements *via* neuronal circuits was laid in evolution. In principle, similar objectives also apply to schistosomes. From the male's perspective, the questions are (i) how does it behave within the final host perceiving its environment to organize migration to target locations such as the portal system of the liver, (ii) how does it find its mating partner, (iii) how does it "learn" from a first pairing experience (capacitation and gaining competence), (iv) how does it move from the liver further on to the mesenteric veins in the gut area while carrying its mate inside the gynecophoral canal, and (v) how does the male "communicate" with its partner during this process and later on after reaching the final destination to organize large-scale egg production and longevity? It appears obvious that molecular communication at different parallel levels is part of the answer, and all transcriptomics data obtained thus far are in favor of this assumption. Data analyses have indicated among others that regulatory RNAs (Cai et al., 2016; Zhu et al., 2016) and kinase activity (Grevelding et al., 2018) but also neuronal regulation are involved (Cai et al., 2016; Wang et al., 2017; Hahnel et al., 2018). In an RNA-seq analysis of *S. mansoni* (Lu et al., 2016; Lu et al., 2017), 39 genes with potential function in neuronal processes (Berriman et al., 2009) were identified to be transcribed in the adult stage with varying transcript levels in whole worms as well as gonads and, to a large extend, in a pairing-dependent manner (**Figure 1A**). Many of these genes were found to be preferentially transcribed in bM and sM but also in unpaired females (sF, single-sex females)

(Lu et al., 2016). Remarkably, the transcript levels of 64% of these genes decreased in mature females (bF, bisex females) after pairing. Similarly, transcripts of some genes detected in ovaries (bO, ovaries of bF; sO, ovaries of sF) and testes (bT, testes of bM; sT, testes of sM) occurred in a tissue-preferential and/or pairing-dependent manner. This included genes with functions in neuronal stem cells such as (i) an ortholog of IRX6 (Smp_149230), which is a homeodomain transcription factor of the iroquois family known to regulate interneuron development (Star et al., 2012) as well as germ-cell maturation in gonads (Kim et al., 2011), and (ii) a neuroglian ortholog (Smp_176350), possibly involved in regulating neuronal circuits (Boerner and Godenschwege, 2010). Of these, a neuroglian ortholog was also listed as a gene showing male-biased transcript occurrence in *S. japonicum* (Cai et al., 2016).

Additional support was obtained from studies about neuropeptidergic signaling, which was discussed playing fundamental roles in flatworm locomotion, feeding, host finding, regeneration, and reproduction (McVeigh et al., 2005; McVeigh et al., 2009; Collins et al., 2010). In an *in silico* analysis, 46 potential flatworm neuropeptides emerging from 32 neuropeptide precursors (npps) have been predicted (Koziol et al., 2016). Of these, transcripts of 7 *npp* genes were localized in the protoscolex of *Echinococcus multilocularis* and appeared to be expressed in the nervous system. In another study, RNA-seq data of the schistosome esophagus area showed that transcripts of one of the *npps* (*Sm_npp_*20a; Smp_088360 with transcripts enriched in males and unpaired females but not in gonads; **Figure 1B**) are enriched in the head part of male worms (Wilson et al., 2015). Furthermore, transcriptional

FIGURE 1 | Heatmaps showing the relative expression of genes annotated as neural genes **(A)** (Berriman et al., 2009, Supplementary Table 9: (https://media.nature.com/original/nature-assets/nature/journal/v460/n7253/extref/nature08160-s2.xls)) and genes potentially coding for neuropeptide precursors **(B)** (Koziol et al., 2016, Supplementary Data S3: (https://www.sciencedirect.com/science/article/abs/pii/S0020751916301217?via%3Dihub)) in *S. mansoni* adults and gonads (red, up-regulation; blue, down-regulation). sM, unpaired males; bM, paired males; sF, unpaired females; bF, paired females; sT, testes of sM; bT, testes of bM; sO, ovaries of sF; bO, ovaries of bF.

profiling comparing *S. mansoni* female head and tail showed that 23 of 27 listed *Sm_npps* expressed in adult *S. mansoni* (**Figure 1B**) preferentially localized in the head part (Wang and Collins, 2016). Therefore, we assume that these NPPs are enriched in the head, possibly in the nervous system. Transcript levels of almost all 27 *Sm_npps* dominated in bM, sM, and sF (**Figure 1B**). Although transcript levels may not be representative for protein levels and/or protein half time, and although also a low amount of protein can be of high cell biological and/or physiological importance, it is tempting to speculate that the comparatively reduced *Sm_npp* transcript levels in bF may point to a lower importance of these neuropeptides (and associated neuronal processes) for females after pairing. In contrast, females before pairing may have different physiological requirements comprising more neuronal and further processes. This view is in line with previous studies concluding that, from a transcriptomic point of view, schistosome females express divergent gene repertoires regulated by pairing (Fitzpatrick and Hoffmann, 2006) and unpaired females are much closer related to males than to paired females (Lu et al., 2016; Grevelding et al., 2018). In view of the male–female interaction of *S. mansoni*, these results also suggest that neuropeptide-mediated regulation circuits are more active in males than in paired females, which may point to a higher importance of neuronal processes for males. Interestingly, a similar tendency with respect to *npp* gene transcription was found in a study about male and female *S. japonicum* (Wang et al., 2017). Among others, the authors investigated different time points throughout the sexual developmental, from pairing to maturation. This included day 16 after final-host infection, when pairing starts, to day 28, when paired females produce eggs. Based on a search for *npp* orthologs of *S. japonicum* in WormBase Parasite (https://parasite.wormbase.org) and looking for their transcript profiles within the data set provided by Wang et al. (2017), a clear tendency can be registered for a reduction of *Sj_npp* transcript levels in females after pairing from day 18 on. In contrast, the transcript levels of these *Sj_npps* remained at a constant level or, in one case (Sjp_0097680, potential ortholog of Smp_052880), increasing level from day 18 on (**Supplementary Data 1**). Figuratively, one could think that schistosome females "hand over" the responsibility for maintaining most of the neuronal circuits involving neuropeptide signaling to their partners after establishing the pairing contact. Indirect support for this assumption was obtained by the analysis of *S. mansoni* G protein-coupled receptors (GPCRs), of which some might represent *Sm_npp* targets. Based on the RNA-seq data, a comparative analysis of the GPCR*ome* generally revealed a pattern of transcriptional activity for the majority of the investigated GPCRs that resembled the patterns of the majority of *Sm_npps*: compared to bF, transcripts of these GPCRs occurred in a higher abundance in sM, bM, and sF (Hahnel et al., 2018). Preliminary data of a deorphanization approach to uncover GPCR-neuropeptide interaction are in support of the assumption that some of the *Sm_npps* and GPCRs correspondingly regulated in the mentioned way may indeed interact (Weth et al., *in preparation*).

Finally, a detailed analysis of transcript levels of two additional genes involved in neuronal processes fits into this scenario. These genes are *S. mansoni* orthologs of a dopa decarboxylase/tyrosine decarboxylases (*Sm-tdc*, Smp_135230), involved in neurotransmitter metabolism (De Luca et al., 2003), and *ebony*

(*Sm-ebony*, Smp_158480), a gene controlling neurotransmitter inactivation (Richardt et al., 2003; Hartwig et al., 2014). Transcript profiling by RNA-seq showed high transcript levels of *Sm-tdc* in bM compared to sM and no transcripts in bF or sF. The meta-analysis of all life stages showed that *Sm-tdc* is stage- and gender-specific expressed in males, being significantly up-regulated after pairing (Lu and Berriman, 2018). Similarly, the amount of *Sm-ebony* transcripts dominated in bM but also in sF, whereas sM and bF showed significantly reduced transcript levels. The meta-analysis showed stage-specific expression in adults with a significant up-regulation in bM and sF (Lu and Berriman, 2018). A comprehensive quantitative reverse transcription-PCR analysis with RNA from males after pairing, separation, and re-pairing *in vitro* finally confirmed the direct influence of pairing on the transcript levels of both genes with a clear bias toward bM (Haeberlein et al., 2019). With respect to *Sm-tdc* and *Sm-ebony*, similar pairing-influenced transcript patterns were found for the orthologs, *SjAADC* and *Sj-ebony*, of *S. japonicum*. Wang et al. (2017) observed an increase of *SjAADC* (Sjp_0075370) and *Sj-ebony* (Sjp_0068110) transcripts in males after pairing, whereas in females transcript levels of these genes remained constant at a low level after pairing. The localization of *SjAADC* transcripts at the gynecophoral canal region (Wang et al., 2017), the interface between paired male and female schistosomes, further supports the view that neuronal processes may govern at least part of the male-female interaction.
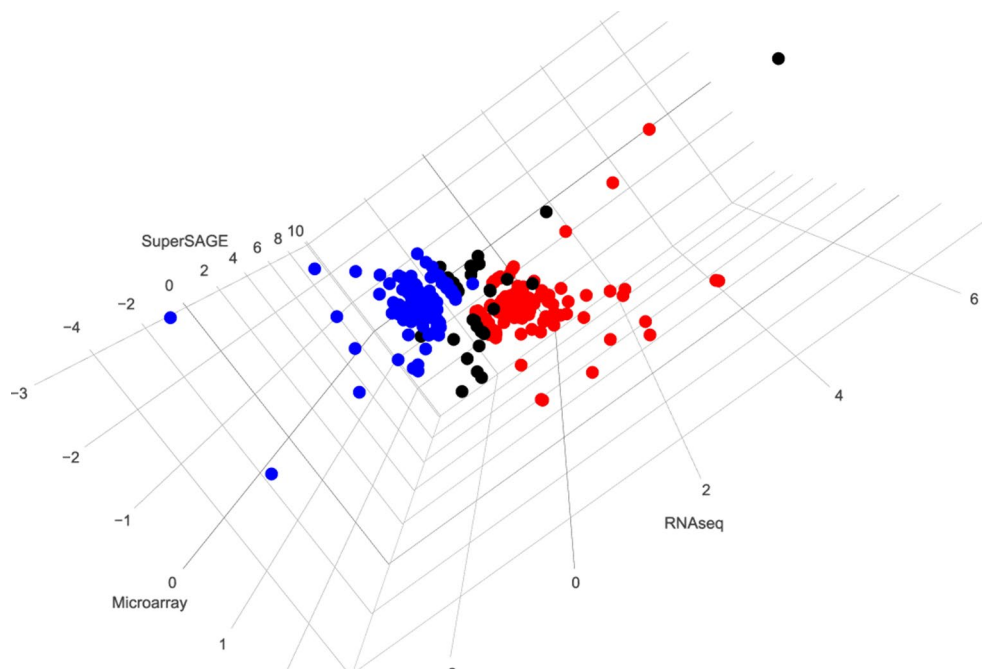
OVERVIEW OF GENES DIFFERENTIALLY TRANSCRIBED IN PAIRED VERSUS UNPAIRED MALES

Based on the existence of three transcriptomics data sets for schistosome males, an overview was generated about all genes commonly found to be significantly differentially transcribed between bM and sM. These data sets were independently obtained by different methods with varying pros and cons depending on the technical basis of these methods (microarray, SuperSAGE, and RNA-seq; for details, see Leutner et al., 2013; Lu et al., 2016; Lu et al., 2017). However, these data sets were produced with RNA of the same origin, a Liberian strain of *S. mansoni* (Grevelding, 1995).

After merging these three data sets, transcripts of 5352 genes were detected by all approaches. Applying the same significance cutoff values used before in these studies (RNA-seq FDR < 0.05; microarray q < 0.01; SuperSAGE p < 1e-10), 154 genes were found to be significantly up-regulated and 153 genes significantly down-regulated after pairing as identified by at least two approaches (**Figure 2** and **Supplementary Table 1**). In particular, we identified 43 genes that were up-regulated (21; bM > sM) or down-regulated (22; bM < sM) in all three data sets, including follistatin (highlighted in **Supplementary Table 1**; sheet 2, no. 21).

CONCLUSION AND PERSPECTIVE

The analysis of transcriptomic data obtained thus far has provided conclusive evidence for a substantial molecular

FIGURE 2 | 3D visualization of genes whose expression is significantly regulated by pairing in at least two data sets. Red (154) and blue (153) dots indicate genes found to be up- or down-regulated, respectively, after pairing as identified by at least two approaches. Black dots (38) indicate different regulation patterns in three data sets. To generate this overview, data from microarray, SuperSAGE analyses (Leutner et al., 2013), and RNA-seq (Lu et al., 2016) were extracted from the respective supplementary tables. To simplify the comparison process, representative gene identifiers were used for all three data sets. In case of multiple-transcript genes, the values were compared among the data sets before one representative was chosen. Significantly differentially regulated genes were selected using the same threshold used in the mentioned studies: q < 0.01 for microarray, p < 1e-10 for SuperSAGE, and FDR < 0.05 for RNA-seq. Transcripts meeting the criteria of occurrence in at least two approaches were picked, and the corresponding genes were visualized using the R package plotly (https://plot.ly) [(**Supplementary Table 1**: http://schisto.xyz/male-deg-3d/ [Interactive visualization of differentially expressed genes between bM and sM from three approaches. The users can use the mouse to zoom in/out and rotate the plot as well as to see detailed information for each gene.)].

contribution of the male to the male–female interaction and the reproductive biology of schistosomes. The male's effort is not restricted to nutritional support and muscle power for carrying the paired female around, which lodges inside the male's gynecophoral canal. Instead, the pairing scenario appears more complex, involving different signaling systems mediating communication between the partners. This includes neuronal processes whose management asymmetrically shifts to the male side upon pairing. Therefore, this review about the males' perspective of the reciprocal sender-recipient relationship of schistosome couples may stimulate future research in this area. Understanding male–female interaction will give a twofold return: (i) for basic science, solving one of the most interesting but yet unanswered question of schistosome biology, and (ii) and for applied research in view of the high demand finding alternative treatment concepts to fight schistosomiasis (Bergquist et al., 2017).

DATA AVAILABILITY

Publicly available datasets were analyzed in this study. This data can be found here: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4976352/

AUTHOR CONTRIBUTIONS

ZL, SS, and OW prepared data and figures and substantially contributed to the work. CG conceived and wrote the manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material (Supplementary Table 1 and Supplementary Data 1) for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00796/full#supplementary-material

REFERENCES

Anderson, L., Amaral, M. S., Beckedorff, F., Silva, L. F., Dazzani, B., Oliveira, K. C., et al. (2015). *Schistosoma mansoni* egg adult male and female comparative gene expression analysis and identification of novel genes by RNA-Seq. *PLoS Negl. Trop. Dis.* 9 (12), e0004334. doi: 10.1371/journal.pntd.0004334

Armstrong, J. C. (1965). Mating behavior and development of schistosomes in the mouse. *J. Parasitol.* 51, 605–616. doi: 10.2307/3276242

Atkinson, K. H., and Atkinson, B. G. (1980). Biochemical basis for the continuous copulation of female *Schistosoma mansoni. Nature* 283 (5746), 478–479. doi: 10.1038/283478a0

Basch, P. F., and Basch, N. (1984). Intergeneric reproductive stimulation and parthenogenesis in *Schistosoma mansoni. Parasitology* 89, 369–376. doi: 10.1017/S0031182000001372

Basch, P. F., and Nicolas, C. (1989). *Schistosoma mansoni*: pairing of male worms with artificial surrogate females. *Exp. Parasitol.* 68 (2), 202–227. doi: 10.1016/0014-4894(89)90098-2

Basch, P. (1991). *Schistosomes: development, reproduction and host relations.* New York, Oxford University Press.

Beckmann, S., Quack, T., Burmeister, C., Buro, C., Long, T., Dissous, C., et al. (2010). *Schistosoma mansoni*: signal transduction processes during the development of the reproductive organs. *Parasitology* 137 (3), 497–520. doi: 10.1017/S0031182010000053

Bergquist, R., Utzinger, J., and Keiser, J. (2017). Controlling schistosomiasis with praziquantel: how much longer without a viable alternative? *Infect. Dis. Poverty* 6 (1), 74. doi: 10.1186/s40249-017-0286-2

Berriman, M., Haas, B. J., LoVerde, P. T., Wilson, R. A., Dillon, G. P., Cerqueira, G. C., et al. (2009). The genome of the blood fluke *schistosoma mansoni. Nature* 460 (7253), 352–358. doi: 10.1038/nature08160

Boerner, J., and Godenschwege, T. A. (2010). Application for the drosophila ventral nerve cord standard in neuronal circuit reconstruction and in-depth analysis of mutant morphology. *J. Neurogenet.* 24 (3), 158–167. doi: 10.3109/01677063.2010.489624

Boissier, J., Grech-Angelini, S., Webster, B. L., Allienne, J. F., Huyse, T., Mas-Coma, S., et al. (2016). Outbreak of urogenital schistosomiasis in Corsica (France): an epidemiological case study. *Lancet Infect. Dis.* 16 (8), 971–979. doi: 10.1016/S1473-3099(16)00175-4

Bostic, J. R., and Strand, M. (1996). Molecular cloning of a *Schistosoma mansoni* protein expressed in the gynecophoral canal of male worms. *Mol. Biochem. Parasitol.* 79 (1), 79–89. doi: 10.1016/0166-6851(96)02640-0

Buro, C., Oliveira, K. C., Lu, Z., Leutner, S., Beckmann, S., and Dissous, C. (2013). Transcriptome analyses of inhibitor-treated schistosome females provide evidence for cooperating Src-kinase and TGFβ receptor pathways controlling mitosis and eggshell formation. *PLoS Pathog.* 9 (6), e1003448. doi: 10.1371/journal.ppat.1003448

Cai, P., Liu, S., Piao, X., Hou, N., Gobert, G. N., McManus, D. P., et al. (2016). Comprehensive transcriptome analysis of sex-biased expressed genes reveals discrete biological and physiological features of male and female *Schistosoma japonicum. PLoS Negl. Trop. Dis.* 10 (4), e0004684. doi: 10.1371/journal.pntd.0004684

Cheng, G., Fu, Z., Lin, J., Shi, Y., Zhou, Y., Jin, Y., et al. (2009). *In vitro* and *in vivo* evaluation of small interference RNA-mediated gynaecophoral canal protein silencing in *Schistosoma japonicum. J. Gene. Med.* 11 (5), 412–421. doi: 10.1002/jgm.1314

Collins, J. J., Hou, X., Romanova, E. V., Lambrus, B. G., Miller, C. M., Saberi, A., et al. (2010). Genome-wide analyses reveal a role for peptide hormones in planarian germline development. *PLoS Biol.* 8, e1000509. doi: 10.1371/journal.pbio.1002234

Conrford, E. M., and Huot, M. E. (1981). Glucose transfer from male to female schistosomes. *Science* 213 (4513), 1269–1271. doi: 10.1126/science.7268436

Cornford, E. M., and Fitzpatrick, A. M. (1985). The mechanism and rate of glucose transfer from male to female schistosomes. *Mol. Biochem. Parasitol.* 17 (2), 131–141. doi: 10.1016/0166-6851(85)90012-X

De Luca, M., Roshina, N. V., Geiger-Thornsberry, G. L., Lyman, R. F., Pasyukova, E. G., and Mackay, T. F. (2003). Dopa decarboxylase (Ddc) affects

variation in *Drosophila* longevity. *Nat. Genet.* 34 (4), 429–433. doi: 10.1038/ng1218

Den Hollander, J. E., and Erasmus, D. A. (1985). *Schistosoma mansoni*: male stimulation and DNA synthesis by the female. *Parasitology* 91, 449–457. doi: 10.1017/S0031182000062697

Erasmus, D. A. (1973). A comparative study of the reproductive system of mature, immature and "unisexual" female *schistosoma mansoni. Parasitology* 67, 165–183. doi: 10.1017/S0031182000046394

Fitzpatrick, J. M., and Hoffmann, K. F. (2006). Dioecious *schistosoma mansoni* express divergent gene repertoires regulated by pairing. *Int. J. Parasitol.* 36, 1081–1089. doi: 10.1016/j.ijpara.2006.06.007

Garchitorena, A., Sokolow, S. H., Roche, B., Ngonghala, C. N., Jocque, M., Lund, A., et al. (2017). Disease ecology, health and the environment: a framework to account for ecological and socio-economic drivers in the control of neglected tropical diseases. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 372 (1722), pii: 20160128. doi: 10.1098/rstb.2016.0128

Giera, M., Kaisar, M. M. M., Derks, R. J. E., Steenvoorden, E., Kruize, Y. C. M., Hokke, C. H., et al. (2018). The *schistosoma mansoni* lipidome: leads for immunomodulation. *Anal. Chim. Acta* 1037, 107–118. doi: 10.1016/j.aca.2017.11.058

Grevelding, C. G. (1995). The female-specific W1 sequence of the Puerto Rican strain of *Schistosoma mansoni* occurs in both genders of a Liberian strain. *Mol. Biochem. Parasitol.* 71 (2), 269–272. doi: 10.1016/0166-6851(94)00058-U

Grevelding, C. G., Sommer, G., and Kunz, W. (1997). Female-specific gene expression in *schistosoma mansoni* is regulated by pairing. *Parasitology* 115, 635–640. doi: 10.1017/S0031182097001728

Grevelding, C. G. (2004). Schistosoma. *Curr. Biol.* 14, R545. doi: 10.1016/j.cub.2004.07.006

Grevelding, C. G., Langner, S., and Dissous, C. (2018). Kinases: molecular stage directors for schistosome development and differentiation. *Trends Parasitol.* 34 (3), 246–260. doi: 10.1016/j.pt.2017.12.001

Gupta, B. C., and Basch, P. F. (1987). Evidence for transfer of a glycoprotein from male to female *schistosoma mansoni* during pairing. *J. Parasitol.* 73 (3), 674–675. doi: 10.2307/3282159

Haeberlein, S., Angrisano, A., Quack, T., Lu, Z., Kellershohn, J., Blohm, A., et al. (2019). *Int. J. Parasitol.* 49 (8), 615–624. doi: 10.1016/j.ijpara.2019.01.006

Hahnel, S., Wheeler, N., Lu, Z., Wangwiwatsin, A., McVeigh, P., Maule, A., et al. (2018). Tissue-specific transcriptome analyses provide new insights into GPCR signalling in adult *schistosoma mansoni. PLoS Pathog.* 14 (1), e1006718. doi: 10.1371/journal.ppat.1006718

Hartwig, S., Dovengerds, C., Herrmann, C., and Hovemann, B. T. (2014). *Drosophila* ebony: a novel type of nonribosomal peptide synthetase related enzyme with unusually fast peptide bond formation kinetics. *FEBS J.* 281 (22), 5147–5158. doi: 10.1111/febs.13054

Haseeb, M. A., Eveland, L. K., and Fried, B. (1985). The uptake, localization and transfer of [4-14C] cholesterol in *schistosoma mansoni* males and females maintained *in vitro. Comp. Biochem. Physiol. A. Comp. Physiol.* 82 (2), 421–423. doi: 10.1016/0300-9629(85)90877-1

Heldin, C. H., and Moustakas, A. (2016). Signaling receptors for TGF-β family members. *Cold Spring Harb. Perspect. Biol.* 8 (8), pii: a022053. doi: 10.1101/cshperspect.a022053

Hoffmann, K. F. (2004). An historical and genomic view of schistosome conjugal biology with emphasis on sex-specific gene expression. *Parasitology* 128 Suppl 1, S11–S22. doi: 10.1017/S0031182004006213

Hu, W., Yan, Q., Shen, D. K., Liu, F., Zhu, Z. D., Song, H. D., et al. (2003). Evolutionary and biomedical implications of a *schistosoma japonicum* complementary DNA resource. *Nat. Genet.* 35, 139–147. doi: 10.1038/ng1236

Kim, B., Kim, Y., Cooke, P. S., Rüther, U., and Jorgensen, J. S. (2011) The fused toes locus is essential for somatic-germ cell interactions that foster germ cell maturation in developing gonads in mice. *Biol. Reprod.* 84 (5), 1024–1032. doi: 10.1095/biolreprod.110.088559

Knobloch, J., Kunz, W., and Grevelding, C. G. (2002). Quantification of DNA synthesis in multicellular organisms by a combined DAPI and BrdU

technique. *Dev. Growth Diff.* 44, 559–563. doi: 10.1046/j.1440-169X.2002.00667.x

Koziol, U., Koziol, M., Preza, M., Costábile, A., Brehm, K., and Castillo, E. (2016). De novo discovery of neuropeptides in the genomes of parasitic flatworms using a novel comparative approach. *Int. J. Parasitol.* 46 (11), 709–721. doi: 10.1016/j.ijpara.2016.05.007

Leutner, S., Oliveira, K. C., Rotter, B., Beckmann, S., Buro, C., Hahnel, S., et al. (2013). Combinatory microarray and SuperSAGE analyses identify pairing-dependently transcribed genes in *schistosoma mansoni* males, including follistatin. *PLoS Negl. Trop. Dis.* 7 (11), e2532. doi: 10.1371/journal.pntd.0002532

LoVerde, P. T., and Chen, L. (1991). Schistosome female reproductive development. *Parasitol. Today* 7 (11), 303–308. doi: 10.1016/0169-4758(91)90263-N

LoVerde, P. T., Osman, A., and Hinck, A. (2007). *Schistosoma mansoni*: TGF-beta signaling pathways. *Exp. Parasitol.* 117 (3), 304–317. doi: 10.1016/j.exppara.2007.06.002

LoVerde, P. T., Andrade, L. F., and Oliveira, G. (2009). Signal transduction regulates schistosome reproductive biology. *Curr. Opin. Microbiol.* 12 (4), 422–428. doi: 10.1016/j.mib.2009.06.005

Lu, Z., Sessler, F., Holroyd, N., Hahnel, S., Quack, T., Berriman, M., et al. (2016). Schistosome sex matters: a deep view into gonad-specific and pairing-dependent transcriptomes reveals a complex gender interplay. *Sci. Rep.* 6, 31150. doi: 10.1038/srep31150

Lu, Z., Sessler, F., Holroyd, N., Hahnel, S., Quack, T., Berriman, M., et al. (2017). A gene expression atlas of adult *Schistosoma mansoni* and their gonads. *Sci. Data* 4, 170118. doi: 10.1038/sdata.2017.118

Lu, Z., and Berriman, M. (2018). Meta-analysis of RNA-seq studies reveals genes responsible for life stage-dominant functions in *Schistosoma mansoni. BioRxiv* 308189, 1–24. doi: 10.1101/308189

Kunz, W. (2001). Schistosome male-female interaction: induction of germ-cell differentiation. *Trends Parasitol.* 17 (5), 227–231. doi: 10.1016/S1471-4922(01)01893-1

Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M., and Gilad, Y. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* 18, 1509–1517. doi: 10.1101/gr.079558.108

Massagué, J., and Chen, Y. G. (2000). Controlling TGF-beta signaling. *Genes Dev.* 14, 627–644. doi: 10.1101/gad.14.6.627

McManus, D. P., Dunne, D. W., Sacko, M., Utzinger, J., Vennervald, B. J., and Zhou, X. N. (2018). Schistosomiasis. *Nat. Rev. Dis. Primers.* 4 (1), 13. doi: 10.1038/s41572-018-0013-8

McVeigh, P., Kimber, M. J., Novozhilova, E., and Day, T. A. (2005). Neuropeptide signalling systems in flatworms. *Parasitology* 131 Suppl, S41–S55. doi: 10.1017/S0031182005008851

McVeigh, P., Mair, G. R., Atkinson, L., Ladurner, P., Zamanian, M., Novozhilova, E., et al. (2009). Discovery of multiple neuropeptide families in the phylum platyhelminthes. *Int. J. Parasitol.* 39 (11), 1243–1252. doi: 10.1016/j.ijpara.2009.03.005

Michaels, R. M. (1969). Mating of *Schistosoma mansoni in vitro. Exp. Parasitol.* 25 (1), 58–71. doi: 10.1016/0014-4894(69)90052-6

Moustakas, A., and Heldin, C. H. (2009). The regulation of TGFbeta signal transduction. *Development* 136, 3699–3714. doi: 10.1242/dev.030338

National Research Council (US) Committee on Research Opportunities in Biology (1989). *Board on Biology, Commission on life sciences, National Research Council.* Raven PH chairman., editor. Washington (DC): The National Academies Press (US). doi: 10.17226/742

Neves, R. H., de Lamare Biolchini, C., Machado-Silva, J. R., Carvalho, J. J., Branquinho, T. B., Lenzi, H. L., et al. (2005). A new description of the reproductive system of *schistosoma mansoni* (Trematoda: Schistosomatidae) analyzed by confocal laser scanning microscopy. *Parasitol. Res.* 95 (1), 43–99. doi: 10.1007/s00436-004-1241-2

Olveda, D. U., Olveda, R. M., McManus, D. P., Cai, P., Chau, T. N., Lam, A. K., et al. (2014). The chronic enteropathogenic disease schistosomiasis. *Infect. Dis.* 28, 193–203. doi: 10.1016/j.ijid.2014.07.009

Osman, A., Niles, E. G., Verjovski-Almeida, S., and LoVerde, P. T. (2006). *Schistosoma mansoni* TGF-beta receptor II: role in host ligand-induced regulation of a schistosome target gene. *PLoS Pathog.* 2 (6), e54. doi: 10.1371/journal.ppat.0020054

Popiel, I., and Basch, P. F. (1984a). Reproductive development of female *Schistosoma mansoni* (Digenea: Schistosomatidae) following bisexual pairing of worms and worm segments. *J. Exp. Zool.* 232 (1), 141–150. doi: 10.1002/jez.1402320117

Popiel, I., and Basch, P. F. (1984b). Putative polypeptide transfer from male to female *Schistosoma mansoni. Mol. Biochem. Parasitol.* 11, 179–188. doi: 10.1016/0166-6851(84)90064-1

Protasio, A. V., Tsai, I. J., Babbage, A., Nichol, S., Hunt, M., Aslett, M. A., et al. (2012). A systematically improved high quality genome and transcriptome of the human blood fluke *schistosoma mansoni. PLoS Negl. Trop. Dis.* 6 (1), e1455. doi: 10.1371/journal.pntd.0001455

Richardt, A., Kemme, T., Wagner, S., Schwarzer, D., Marahiel, M. A., and Hovemann. B. T. (2003). Ebony, a novel nonribosomal peptide synthetase for beta-alanine conjugation with biogenic amines in Drosophila. *J. Biol. Chem.* 278 (42), 41160–41166. doi: 10.1074/jbc.M304303200

Ruppel, A., and Cioli, D. (1977). A comparative analysis of various developmental stages of *schistosoma mansoni* with respect to their protein composition. *Parasitology* 75 (3), 339–343. doi: 10.1017/S003118200005188X

Schistosoma japonicum Genome Sequencing and Functional Analysis Consortium (2009). The *Schistosoma japonicum* genome reveals features of host-parasite interplay. *Nature* 460 (7253), 345–351. doi: 10.1038/nature08140

Shaw, J. R., Marshall, I., and Erasmus, D. A. (1977). *Schistosoma mansoni: in vitro* stimulation of vitelline cell development by extracts of male worms. *Exp. Parasitol.* 42 (1), 14–20. doi: 10.1016/0014-4894(77)90056-X

Shaw, M. K. (1987). *Schistosoma mansoni*: vitelline gland development in females from single sex infections. *J. Helminthol.* 61, 253–259. doi: 10.1017/S0022149X00010117

Silveira, A. M., Friche, A. A., and Rumjanek, F. D. (1986). Transfer of [14C] cholesterol and its metabolites between adult male and female worms of *schistosoma mansoni. Comp. Biochem. Physiol. B.* 85 (4), 851–857. doi: 10.1016/0305-0491(86)90186-0

Smit, C. H., van Diepen, A., Nguyen, D. L., Wuhrer, M., Hoffmann, K. F., Deelder, A. M., et al. (2015). Glycomic analysis of life stages of the human parasite *schistosoma mansoni* reveals developmental expression profiles of functional and antigenic glycan motifs. *Mol. Cell. Proteomics* 14 (7), 1750–1769. doi: 10.1074/mcp.M115.048280

Sotillo, J., Doolan, D., and Loukas, A. (2017). Recent advances in proteomic applications for schistosomiasis research: potential clinical impact. *Expert Rev. Proteomics* 14 (2), 171–183. doi: 10.1080/14789450.2017.1271327

Standley, C. J., Mugisha, L., Dobson, A. P., and Stothard, J. R. (2012). Zoonotic schistosomiasis in non-human primates: past, present and future activities at the human-wildlife interface in Africa. *J. Helminthol.* 86 (2), 131–140. doi: 10.1017/S0022149X12000028

Star, E. N., Zhu, M., Shi, Z., Liu, H., Pashmforoush, M., Sauve, Y., et al. (2012). Regulation of retinal interneuron subtype identity by the Iroquois homeobox gene Irx6. *Development* 139 (24), 4644–4655. doi: 10.1242/dev.081729

Verjovski-Almeida, S., DeMarco, R., Martins, E. A., Guimarães, P. E., Ojopi, E., Paquola, A. C., et al. (2003). Transcriptome analysis of the acoelomate human parasite *schistosoma mansoni. Nat. Genet.* 35, 148–157. doi: 10.1038/ng1237

Wang, J., and Collins, J. J. 3rd (2016). Identification of new markers for the *schistosoma mansoni* vitelline lineage. *Int. J. Parasitol.* 46 (7), 405–410. doi: 10.1016/j.ijpara.2016.03.004

Wang, J., Yu, Y., Shen, H., Qing, T., Zheng, Y., Li, Q., et al. (2017). Dynamic transcriptomes identify biogenic amines and insect-like hormonal regulation for mediating reproduction in *schistosoma japonicum. Nat. Commun.* 8, 14693. doi: 10.1038/ncomms14693

Waisberg, M., Lobo, F. P., Cerqueira, G. C., Passos, L. K., Carvalho, O. S., Franco, G. R., et al. (2007). Microarray analysis of gene expression induced by sexual contact in *schistosoma mansoni. BMC Genomics* 8, 181. doi: 10.1186/1471-2164-8-181

Weissenbach, J. (2016). The rise of genomics. *C. R. Biol.* 339 (7-8), 231–239. doi: 10.1016/j.crvi.2016.05.002

Williams, D. L., Sayed, A. A., Bernier, J., Birkeland, S. R., Cipriano, M. J., Papa, A. R., et al. (2007). Profiling *schistosoma mansoni* development using serial analysis of gene expression (SAGE). *Exp. Parasitol.* 117 (3), 246–258. doi: 10.1016/j.exppara.2007.05.001

Wilson, R. A., Li, X. H., MacDonald, S., Neves, L. X., Vitoriano-Souza, J., Leite, L. C., et al. (2015). The schistosome esophagus is a 'hotspot' for microexon and

lysosomal hydrolase gene expression: implications for blood processing. *PLoS Negl. Trop. Dis.* 9 (12), e0004272. doi: 10.1371/journal.pntd.0004272

World Health Organization, Schistosomiasis (2019). Available from: https://www.who.int/news-room/fact-sheets/detail/schistosomiasis

Young, N. D., Jex, A. R., Li, B., Liu, S., Yang, L., Xiong, Z., et al. (2012). Whole-genome sequence of *schistosoma haematobium*. *Nat. Genet.* 44, 221–225. doi: 10.1038/ng.1065

Zhu, L., Zhao, J., Wang, J., Hu, C., Peng, J., Luo, R., et al. (2016). MicroRNAs are involved in the regulation of ovary development in the pathogenic blood fluke *schistosoma japonicum*. *PLoS Pathog.* 12 (2), e1005423. doi: 10.1371/journal.ppat.1005423

Weighted Gene Co-Expression Analyses Point to Long Non-Coding RNA Hub Genes at Different *Schistosoma mansoni* Life-Cycle Stages

Lucas F. Maciel[1,2], David A. Morales-Vicente[1,3], Gilbert O. Silveira[1,3], Raphael O. Ribeiro[1,3], Giovanna G. O. Olberg[1], David S. Pires[1], Murilo S. Amaral[1] and Sergio Verjovski-Almeida[1,3]*

[1] Laboratório de Expressão Gênica em Eucariotos, Instituto Butantan, São Paulo, Brazil, [2] Programa Interunidades em Bioinformática, Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, Brazil, [3] Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil

Long non-coding RNAs (lncRNAs) (>200 nt) are expressed at levels lower than those of the protein-coding mRNAs, and in all eukaryotic model species where they have been characterized, they are transcribed from thousands of different genomic *loci*. In humans, some four dozen lncRNAs have been studied in detail, and they have been shown to play important roles in transcriptional regulation, acting in conjunction with transcription factors and epigenetic marks to modulate the tissue-type specific programs of transcriptional gene activation and repression. In *Schistosoma mansoni*, around 10,000 lncRNAs have been identified in previous works. However, the limited number of RNA-sequencing (RNA-seq) libraries that had been previously assessed, together with the use of old and incomplete versions of the *S. mansoni* genome and protein-coding transcriptome annotations, have hampered the identification of all lncRNAs expressed in the parasite. Here we have used 633 publicly available *S. mansoni* RNA-seq libraries from whole worms at different stages (n = 121), from isolated tissues (n = 24), from cell-populations (n = 81), and from single-cells (n = 407). We have assembled a set of 16,583 lncRNA transcripts originated from 10,024 genes, of which 11,022 are novel *S. mansoni* lncRNA transcripts, whereas the remaining 5,561 transcripts comprise 120 lncRNAs that are identical to and 5,441 lncRNAs that have gene overlap with *S. mansoni* lncRNAs already reported in previous works. Most importantly, our more stringent assembly and filtering pipeline has identified and removed a set of 4,293 lncRNA transcripts from previous publications that were in fact derived from partially processed mRNAs with intron retention. We have used weighted gene co-expression network analyses and identified 15 different gene co-expression modules. Each parasite life-cycle stage has at least one highly correlated gene co-expression module, and each module is comprised of hundreds to thousands lncRNAs and mRNAs having correlated co-expression patterns at different stages. Inspection of the top most

highly connected genes within the modules' networks has shown that different lncRNAs are hub genes at different life-cycle stages, being among the most promising candidate lncRNAs to be further explored for functional characterization.

INTRODUCTION

Schistosomiasis is a neglected tropical disease, caused by flatworms from the genus *Schistosoma*, with estimates of more than 250 million infected people worldwide and responsible for 200 thousand deaths annually at the Sub-Saharan Africa (Who, 2015). *Schistosoma mansoni*, prevalent in Africa and Latin America, is one of the three main species related to human infections (Cdc, 2018). In America, it is estimated that 1 to 3 million people are infected by *S. mansoni* and over 25 million live in risk areas, being Brazil and Venezuela the most affected (Zoni et al., 2016). The prevalence of this disease is correlated to social–economic and environmental factors (Gomes Casavechia et al., 2018).

This parasite has a very complex life-cycle comprised of several developmental stages, with a freshwater snail intermediate-host and a final mammalian host (Basch, 1976). Recently, it has been shown that epigenetic changes are required for life-cycle progression (Roquis et al., 2018). However, little is known about the genes and molecules that drive this process through the life-cycle stages of *S. mansoni*. A better understanding of the gene expression regulation mechanisms and of their components may lead to new therapeutic targets (Batugedara et al., 2017), and one key element could be the long non-coding RNAs (lncRNAs) (Blokhin et al., 2018).

LncRNAs are defined as transcripts longer than 200 nucleotides, without apparent protein-coding potential (Cao et al., 2018). The term "apparent" is included because it is already known that some lncRNAs actually have dual function roles, being functional both as lncRNAs and through peptides shorter than 100 amino acids that they encode (Nam et al., 2016; Choi et al., 2018). In mammalians, lncRNAs regulate gene expression through different mechanisms (Bhat et al., 2016), including mediating epigenetic modifications (Hanly et al., 2018), and were shown to be important in vital processes, such as cell cycle regulation (Kitagawa et al., 2013), pluripotency maintenance (Rosa and Ballarino, 2016), and reproduction (Golicz et al., 2018).

In *S. mansoni*, the expression of lncRNAs at different life-cycle stages was first detected by our group in 2011 using microarrays (Oliveira et al., 2011). Subsequently, large-scale identification of *S. mansoni* lncRNAs has been reported in three studies from our group and from others that analyzed high-throughput RNA-sequencing (RNA-seq) data (Vasconcelos et al., 2017; Liao et al., 2018; Oliveira et al., 2018), but each of them has used a the limited number of data sets (from 4 to 88 RNA-seq libraries). Because each work used different mapping tools and parameters (Vasconcelos et al., 2017; Liao et al., 2018;

Oliveira et al., 2018), and given that Liao et al. (2018) did not compare their lncRNAs with the previously published ones, part of the lncRNAs are redundant among the three reports. In addition, the lncRNAs were annotated against the old version 5.2 of the genome and protein-coding transcriptome (Protasio et al., 2012); as a result, a set of transcripts that were previously annotated as lncRNAs (Vasconcelos et al., 2017; Liao et al., 2018; Oliveira et al., 2018), seem now to represent partially processed pre-mRNAs arising from novel protein-coding genes annotated in the new version 7.1 of the transcriptome (https://parasite.wormbase.org/Schistosoma_mansoni_prjea36577/); these transcripts were previously annotated as having no coding potential due to intron retention, as exemplified in **Supplementary Figure S1**. Besides, these three works used expression data from whole parasites, while it is known from other species that lncRNAs have tissue- and cell-specific expression (Wu et al., 2016; Credendino et al., 2017).

The aim of the present work is to identify and annotate a robust and more complete set of lncRNAs that agrees with the most updated transcriptome annotation, and to analyze RNA-seq data sets still non-annotated for the presence of lncRNAs—e.g., gonads (Lu et al., 2016) and single-cell (Tarashansky et al., 2018; Wang et al., 2018) RNA-seq libraries. The goal is to provide a foundation that will enable future studies on the role of lncRNAs in *S. mansoni* biology, which could eventually identify potential new therapeutic targets.

MATERIALS AND METHODS

Transcripts Reconstruction

To identify new lncRNAs, 633 publicly available RNA-seq libraries from whole worms at different stages (miracidia, n = 1; sporocysts, n = 1; cercariae, n = 8; schistosomula, n = 11; juveniles, n = 9; adult males, n = 34; adult females, n = 37; and mixed adults, n = 20), from tissues (testes, n = 6; ovaries, n = 5; posterior somatic tissues, n = 3; heads, n = 5; and tails, n = 5), from cell populations (n = 81) and from single cells (from juveniles, n = 370 and mother sporocysts stem cells, n = 37) were downloaded from the SRA and ENA databases (**Supplementary Table S1**). The only whole-worm stage that was not included was eggs, because there is a single RNA-seq library available in the public domain (Anderson et al., 2016), which has only 252,000 egg reads, an amount that is fourfold lower than the minimum number of reads per library in the other whole-worm libraries that we used (namely 1 million good quality reads), being a too-low coverage for an unbiased detection of stage- or tissue-specific lncRNAs in complex

organisms (Sims et al., 2014). The new versions of the genome (v 7) and transcriptome (v 7.1), which were used as reference in this study, were downloaded from the WormBase ParaSite resource (Howe et al., 2017) at https://parasite.wormbase.org/Schistosoma_mansoni_prjea36577/.

Quality control was done with fastp v 0.19.4 (Chen et al., 2018) (default parameters), removing adapters and low-quality reads. The reads in each library were then mapped against the genome with STAR v 2.6.1c in a two-pass mode, with parameters indicated by STAR's manual as the best ones to identify new splicing sites and transcripts (Dobin et al., 2013). RSeQC v 2.6.5 (Wang et al., 2012) was used to identify RNA-Seq library strandedness to be used in transcripts reconstruction and expression levels quantification. For each library, multi mapped reads were removed with Samtools v 1.3 (Li et al., 2009) and uniquely mapped reads were used for transcript reconstruction with Scallop v 10.2 (–min_mapping_quality 255 -min_splice_ boundary_hits 2) (Shao and Kingsford, 2017). A new splicing site should be confirmed at least by two reads to be considered. A consensus transcriptome from all libraries was built using TACO v 0.7.3 (–filter-min-length 200 -isoform-frac 0.05), an algorithm that reconstructs the consensus transcriptome from a collection of individual assemblies (Niknafs et al., 2016). As described by Niknafs et al. (2016), TACO employs change point detection to break apart complex loci and correctly delineate transcript start and end sites and a dynamic programming approach to assemble transcripts from a network of splicing patterns (Niknafs et al., 2016).

LncRNAs Classification

In the consensus transcriptome, transcripts shorter than 200 nt, monoexonic or with exon-exon overlap with protein-coding genes from the same genomic strand were removed from the set. The coding potential of the remaining transcripts was evaluated by means of the FEELnc tool v 0.1.1 (Wucher et al., 2017) with shuffle mode, which uses a random forest machine-learning algorithm and classifies these transcripts into lncRNAs or protein-coding genes, and also by CPC2 v 0.1 (Yang et al., 2017), which classifies through a support vector machine model using four intrinsic features. Only transcripts classified as lncRNAs by both tools were kept. ORFfinder v 0.4.3 (https://www.ncbi.nlm.nih.gov/orffinder/) was used to extract the putative longest open reading frames (ORFs); these putative peptides were then submitted to orthology-based annotation with eggNOG-mapper webtool (HMMER mapping mode) (Huerta-Cepas et al., 2017). Transcripts with no hits against the eukaryote eggNOG database were then considered as lncRNAs. If any transcript isoform was classified as a protein-coding mRNA at any step, all transcripts mapping to the same genomic *locus* were removed to avoid eventual pre-mRNAs. After this final step, a lncRNAs GTF file was created.

Histone Marks

To identify histone H3 lysine 4 trimethylation (H3K4me3) and H3 lysine 27 trimethylation (H3K27me3) marks near the

transcription start site (TSS) of lncRNAs, we used 12 libraries of Chromatin Immunoprecipitation Sequencing (ChIP-Seq) data generated by Roquis et al. (2018) for cercariae, schistosomula, and adults (**Supplementary Table S1**), which had more than 90% overall mapping rate. The reads were downloaded from the SRA database and mapped against the genome v 7 with Bowtie2 v 2.3.4.3 (Langmead and Salzberg, 2012) (parameters end-to-end, -sensitive, -gbar 4). Because there are no input data sets publicly available in the SRA database for the Roquis et al. (2018) paper, we were not able to exactly reproduce the pipeline that was described in the Methods section of that paper, which used the input as a reference for peak calling. Instead, we used HOMER v 4.10 (Heinz et al., 2010) for removing multi-mapped and duplicated reads and for significant peak calling as described by Anderson et al. (2016), an approach also used by Vasconcelos et al. (2017) in the first large-scale annotation of lncRNAs in *S. mansoni*. The number of reads in the peak should be at least fourfold higher than in the peaks of the surrounding 10-kb area and the Poisson p-value threshold cutoff was 0.0001. The lncRNAs with significant histone mark peaks within 1-kb distance upstream and downstream from their TSS were annotated. The lncRNAs with overlapping marks are shown with an intersection diagram that was plotted using the UpSetR tool v 1.3.3 (Lex et al., 2014). The Venn diagram tool at http://bioinformatics.psb.ugent.be/beg/tools/venn-diagrams was used for generating the lists of lncRNA genes belonging to each intersection set.

Co-Expression Networks

The lncRNAs GTF file was then added to the *S. mansoni* public protein-coding transcriptome version 7.1 GTF file, and the resulting protein-coding + lncRNAs GTF was used as the reference together with the genome sequence v.7 for mapping the reads of each RNA-seq library under study, again using the STAR tool, now in the one pass mode, followed by gene expression quantification with RSEM v 1.3 (Li and Dewey, 2011). Weighted gene co-expression network analyses v 1.68 (WGCNA) (Langfelder and Horvath, 2008) were then performed to identify modules related to the life-cycle stages and tissues of the organism. For this purpose, only libraries from whole worms or from tissues with more than 50% of the reads uniquely mapped were used. To reduce noise, only transcripts with expression greater than 1 transcript per million (TPM) in at least half of the libraries in one or more stages/tissues were considered. Expression levels were measured in log space with a pseudocount of 1 (log2 (TPM+1)), and we set the transcript expression to zero when log2 (TPM+1) <1. For the construction of the adjacency matrix, the power adjacency function for signed networks was applied with the soft-thresholding beta parameter equal to 14, which resulted in a scale-free topology model fit index ($R^2 = 0.935$). The adjacency matrix was then converted to the Topological Overlap Matrix (TOM) and the dissimilarity TOM (1 – TOM) was calculated (Langfelder and Horvath, 2008).

Correlation between the modules and the stages was calculated based on the Pearson correlation coefficient between

the expression levels of the transcripts belonging to each module along the stages, as suggested in the WGCNA tutorial (https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/Tutorials/). As miracidia and sporocysts have only one library each, are closely related stages of development, and were clustered together as an outgroup based on their overall expression patterns (as shown in the Results), we decided to consider both stages together as one group (miracidia/sporocysts) to calculate the correlation and p-values between modules and stages.

The Gene Trait Significance (GS) was calculated based on the correlation of an individual transcript and the trait, which in our case was always the stage of higher absolute Pearson correlation coefficient with the module where the transcript belongs. For example, for a transcript that belongs to the red module (most highly correlated with testes, see Results), the correlation was calculated between the expression of the transcript in the testes libraries and the expression of the transcript in all other non-testes libraries.

Gene Ontology (GO) Enrichment

Protein-coding genes were submitted to eggNOG-mapper (Huerta-Cepas et al., 2017) for annotation of GO terms. Based on this annotation (available at **Supplementary Table S2**), we performed GO enrichment analyses with BINGO (Maere et al., 2005). For each module, we used a hypergeometric test, the whole annotation as reference set, and FDR ≤ 0.05 was used as the significance threshold.

Single-Cell Analyses

The expression levels were quantified in single-cell RNA-seq libraries from juveniles' stem cells (Tarashansky et al., 2018) and mother sporocysts stem cells (Wang et al., 2018) by RSEM. We used Scater v 1.10.1 (Mccarthy et al., 2017) to normalize and identify high-quality single-cell RNA-Seq libraries, i.e., those that have at least 100,000 total counts and at least 1,000 different expressed transcripts, as recommended by Mccarthy et al. (2017); all libraries were classified as high quality.

Next, we used the R package Single-Cell Consensus Clustering (SC3) tool v 1.10.1 (Kiselev et al., 2017), which performs an unsupervised clustering of scRNA-seq data. Based on the clusters identified, we used the plot SC3 markers function to find marker genes based on the mean cluster expression values. These markers are highly expressed in only one of the clusters and indicate the specific expression at the cell level. As described by Kiselev et al. (2017), the area under the receiver operating characteristic (ROC) curve is used to quantify the accuracy of the prediction. A p-value is assigned to each gene by using the Wilcoxon signed rank test. Genes with the area under the ROC curve (AUROC) > 0.85 and with p-value < 0.01 are defined as marker genes.

Parasite Materials

All parasite materials were from a BH isolate of *S. mansoni* maintained by passage through golden hamster (*Mesocricetus auratus*) and *Biomphalaria glabrata* snails. Eggs were purified from livers of hamsters previously infected with *S. mansoni*, according to Dalton et al. (1997). After purification, eggs were added to 10 ml of distilled water and exposed to a bright light. Supernatant containing hatched miracidia was removed every 30 min for 2 h and replaced by fresh water. The supernatants containing the miracidia were pooled and chilled on ice, and miracidia were then recovered by centrifugation at 15,000g for 20 s (Dalton et al., 1997). Supernatant was discarded and miracidia stored in RNAlater (Ambion) until RNA extraction.

Cercariae were collected from snails infected with 10 miracidia each. Thirty-five days after infection, the snails were placed in the dark in water and then illuminated for 2 h to induce shedding. The emerging cercariae were collected by centrifugation, washed with PBS once, and then stored in RNAlater (Ambion) until RNA extraction.

Schistosomula were obtained by mechanical transformation of cercariae and separation of their bodies as previously described (Basch, 1981), with some modifications. Briefly, cercariae were collected as described above and then suspended in 15 ml of M169 medium (Vitrocell, cat number 00464) containing penicillin/streptomycin, amphotericin (Vitrocell, cat number 00148). Mechanical transformation was performed by passing the cercariae 10 times through a 23G needle. To separate schistosomula from the tails, the tail-rich supernatant was decanted and the sedimented bodies resuspended in a further 7 ml of M169 medium. The procedure was repeated until less than 1% of the tails remained. The newly transformed schistosomula were maintained for 24 h in M169 medium (Vitrocell, cat number 00464) supplemented with penicillin/streptomycin, amphotericin, gentamicin (Vitrocell, cat number 00148), 2% fetal bovine serum, 1 µM serotonin, 0.5 µM hypoxanthine, 1 µM hydrocortisone, and 0.2 µM triiodothyronine at 37°C and 5% CO_2. Schistosomula cultivated for 24 h were collected, washed three times with PBS and stored in RNAlater (Ambion) until RNA extraction.

Adult *S. mansoni* worms were recovered by perfusion of golden hamsters that had been infected with 250 cercariae, 7 weeks previously. Approximately 200 *S. mansoni* (BH strain) adult worm pairs were freshly obtained through the periportal perfusion of hamster, as previously described (Anderson et al., 2016; Vasconcelos et al., 2017). After perfusion, the adult worm pairs were kept for 3 h at 37°C and 5% CO_2 in Advanced RPMI Medium 1640 (Gibco, 12633-012) supplemented with 10% fetal bovine serum, 12 mM HEPES (4-(2-hydroxyethyl) piperazine-1-ethanesulfonic acid) pH 7.4, and 1% penicillin/ streptomycin, amphotericin (Vitrocell, cat number 00148). After 3 h of incubation, the adult worm pairs were collected, washed three times with PBS, and stored in RNAlater (Ambion) until RNA extraction. Before the extraction of RNA from males or females, adult worm pairs were manually separated in RNAlater (Ambion) using tweezers.

RNA Extraction, Quantification, and Quality Assessment

Total RNA from eggs (E), miracidia (Mi), cercariae (C), and schistosomula (S) was extracted according to

Vasconcelos et al. (2017). Briefly, 100,000 eggs, 15,000 miracidia, 25,000 cercariae, or 25,000 schistosomula were ground with glass beads in liquid nitrogen for 5 min. Then, the Qiagen RNeasy Micro Kit (Cat number 74004) was used for RNA extraction and purification according to the manufacturer's instructions, except for the DNase I treatment, the amount of DNase I was doubled and the time of treatment was increased to 45 min.

Male (M) or female (F) adult worms were first disrupted in Qiagen RLT buffer using glass potters and pestles. RNA from males or females was then extracted and purified using the Qiagen RNeasy Mini Kit (Cat number 74104), according to the manufacturer's instructions, except for the DNase I treatment, which was the same used for egg, miracidia, cercariae, and schistosomula RNA extraction.

All the RNA samples were quantified using the Qubit RNA HS Assay Kit (Q32852, Thermo Fisher Scientific), and the integrity of RNAs was verified using the Agilent RNA 6000 Pico Kit (5067-1513 Agilent Technologies) in a 2100 Bioanalyzer Instrument (Agilent Technologies). Four biological replicates were assessed for each life cycle stage, except for schistosomula, for which three biological replicates were assessed.

Reverse Transcription and Quantitative PCR (qPCR) Assays

The reverse transcription (RT) reaction was performed with 200 ng of each total RNA sample using the SuperScript IV First-Strand Synthesis System (18091050; Life Technologies) and random hexamer primers in a 20-μL final volume. The obtained complementary DNAs (cDNAs) were diluted four times in DEPC water, and quantitative PCR was performed using 2.5 μL of each diluted cDNA in a total volume of 10 μL containing 1X LightCycler 480 SYBR Green I Master Mix (04707516001, Roche Diagnostics) and 800 nM of each primer in a LightCycler 480 System (Roche Diagnostics). Primers for selected transcripts (**Supplementary Table S3**) were designed using the Primer 3 tool (http://biotools.umassmed.edu/bioapps/primer3_www.cgi), and each real-time qPCR was run in two technical replicates. The results were analyzed by comparative Ct method (Livak and Schmittgen, 2001). Real-time data were normalized in relation to the level of expression of Smp_090920 and Smp_062630 reference genes.

RESULTS

LncRNAs Identification and Annotation

Using 633 publicly available *S. mansoni* RNA-seq libraries from whole worms at different stages, from isolated tissues, from cell-populations, and from single-cells (see Methods), our pipeline assembled a consensus transcriptome comprised of 78,817 transcripts, of which 7,954 were classified as intergenic lncRNAs (lincRNAs), 7,438 as antisense lncRNAs, and 1,191 as sense lncRNAs, totalizing 16,583 lncRNA transcripts originated from 10,024 genes (on average, 1.65 lncRNA isoforms per lncRNA gene); the summary of all six filtering steps in the pipeline is presented in **Table 1**. With the FEELnc lncRNA classification

tool (**Table 1, step 5**), the most important feature for transcripts classification was the ORF coverage (**Supplementary Figure S2A**), i.e., the fraction of the total length of the transcript that is occupied by the longest predicted ORF. In the FEELnc model training process, an optimal coding probability cutoff (0.348) was identified, which resulted in 0.962 sensitivity and specificity of mRNA classification (**Supplementary Figure S2B**). Analogous information is not provided in the output of the CPC2 classification tool (**Table 1, step 5**). Only the lncRNAs classified as such by both prediction tools were retained in the final set (**Table 1**).

From the total set of 16,583 lncRNAs obtained here, 11,022 are novel *S. mansoni* lncRNAs, whereas the remaining 5,561 transcripts comprise 120 lncRNAs that are identical to previously published ones, and 5,441 lncRNAs that have gene overlap with *S. mansoni* lncRNAs already reported in previous works (Vasconcelos et al., 2017; Liao et al., 2018; Oliveira et al., 2018) (**Supplementary Table S4**). In particular, among the 7,029 lincRNAs previously published ones reported by our group (Vasconcelos et al., 2017), a total of 4,368 transcripts have partial or complete sequence overlap with the lncRNAs obtained here, whereas the remaining 2,661 (37.8%) transcripts previously annotated by Vasconcelos et al. (2017) are no longer in the present updated *S. mansoni* lncRNAs data set.

Among the transcripts in the public data set that were previously classified as lncRNAs (Vasconcelos et al., 2017; Liao et al., 2018; Oliveira et al., 2018) and are now excluded, a total of 4,293 were reconstructed in our assembly; however, they were removed from our set of lncRNAs because they were partially processed pre-mRNA transcripts that have exon-exon overlap with new protein-coding genes of version 7.1. The remaining transcripts previously classified as lncRNAs were reconstructed here but were removed by the more stringent, presently used filtering steps. We have created a track on the *S. mansoni* UCSC-like genome browser (http://schistosoma.usp.br/), where the set

TABLE 1 | Summary of transcripts removed at each filtering step and the final set of *S. mansoni* lncRNAs.

Pipeline step number	Removed transcripts	Remaining transcripts (Genes)
1. Total assembled transcripts		78,817 (42,337)
2. Remove short transcripts (<200 nt)	11	78,806
3. Remove monoexonic transcripts	27,255	51,551
4. Remove transcripts that overlap exon-exon with known Sm protein-coding genes	31,183	20,368
5. Remove transcripts with coding potential (FEELnc and/ or CPC2 tools)	3,618	16,750
6. Remove transcripts with hits on eggNOG-Mapper	167	16,583
7. Total lncRNAs identified		16,583 (10,024)
Long intergenic non-coding RNAs		7,954
Antisense long non-coding RNAs		7,438
Sense long non-coding RNAs		1,191

of 16,583 lncRNAs obtained here can be visualized and the GTF and BED files can be downloaded. In **Figure 1**, we show a selected protein-coding desert genomic *locus* on chromosome 2 covering 245 kilobases, which harbors only three protein-coding genes and where we identified seven lincRNAs, two sense lncRNAs, and one antisense lncRNA that were not previously described.
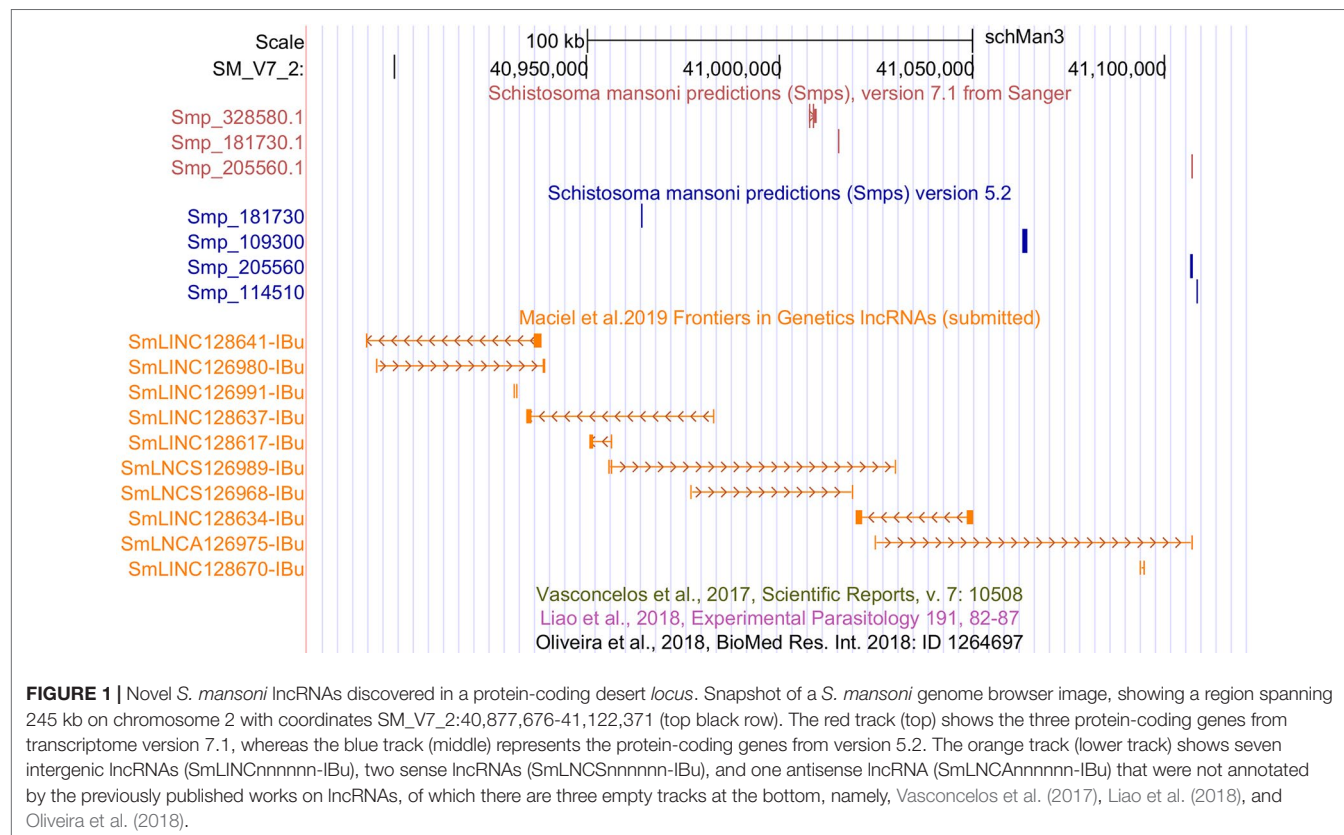
To identify the contribution from each type of RNA-Seq library to the final lncRNAs set, we used the TACO transcriptome assembler to obtain the transcriptomes of the four following groups: whole organisms, tissues, cell populations, and single cells. The result is presented in **Figure 2** and shows that each type of sample contributed with at least 1,000 unique lncRNAs, detected only in that group. It is worthy to mention that around 4% of the 16,583 lncRNAs are lost when the four transcriptomes are reconstructed separately.
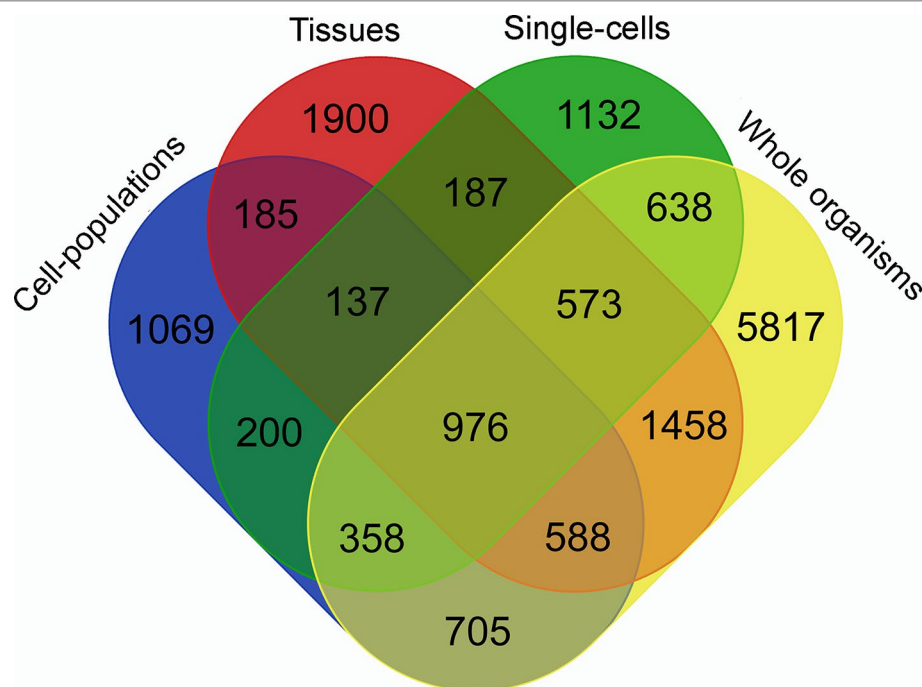
Almost all lncRNAs encode short canonical ORFs within their sequences, however, as described by Verheggen et al. (2017), one can evaluate if these ORFs are originated only by random nucleotide progression by comparing the relative sizes of ORFs using the reverse-complement of the sequence as a control. As presented in **Figure 3**, it is very clear that the size distribution of *bona fide S. mansoni* mRNA ORFs (sense) from the annotated v 7.1 transcriptome is greatly shifted toward longer sizes, compared with the size distribution of random ORFs found in their reverse-complement sequences. It is also possible to observe that the size distribution of ORFs found both within the lncRNAs (sense) and within their

reverse-complement sequences is very similar and is also similar to the size distribution of random ORFs in the reverse-complement sequence of mRNAs.
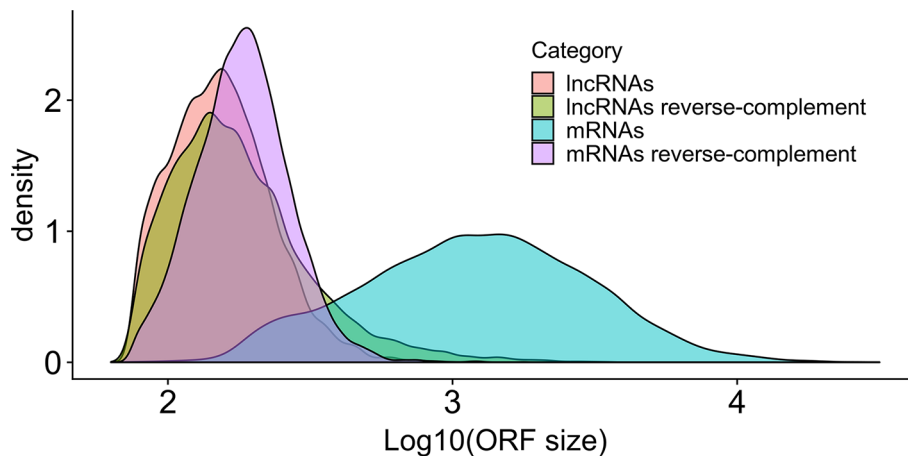
Histone Marks at the TSS of LncRNAs as Evidence of Regulation

As reported earlier, cross-matching of the lncRNAs genomic coordinates with the genomic coordinates of different publicly available histone mark profiles, obtained by ChIP-Seq at different life-cycle stages, adds another layer of functionality evidence for this class of RNAs (Vasconcelos et al., 2017; Cao et al., 2018). We used the data for two different histone marks obtained by Roquis et al. (2018) in cercariae, schistosomula, and adult parasites, namely, H3K4me3 that is generally associated with active transcription, and H3K27me3 associated to transcription repression (Barski et al., 2007). First, we analyzed the histone mark profiles of H3K4me3 and H3K27me3 around the TSS of protein-coding genes through the stages, and they were very similar to the ones presented by Roquis et al. (2018) (**Supplementary Figure S3**). **Figure 4** shows that these marks are also present around the TSS of *S. mansoni* lncRNAs at the three different life-cycle stages; a comparison with **Supplementary Figure S3** shows that these marks are less abundant in lncRNAs than that in the protein-coding genes loci and more spread away of the lncRNAs TSSs when compared with protein-coding genes. This profile is

FIGURE 1 | Novel *S. mansoni* lncRNAs discovered in a protein-coding desert *locus*. Snapshot of a *S. mansoni* genome browser image, showing a region spanning 245 kb on chromosome 2 with coordinates SM_V7_2:40,877,676-41,122,371 (top black row). The red track (top) shows the three protein-coding genes from transcriptome version 7.1, whereas the blue track (middle) represents the protein-coding genes from version 5.2. The orange track (lower track) shows seven intergenic lncRNAs (SmLINCnnnnnn-IBu), two sense lncRNAs (SmLNCSnnnnnn-IBu), and one antisense lncRNA (SmLNCAnnnnnn-IBu) that were not annotated by the previously published works on lncRNAs, of which there are three empty tracks at the bottom, namely, Vasconcelos et al. (2017), Liao et al. (2018), and Oliveira et al. (2018).

FIGURE 2 | Venn diagram representing the specific contribution from each type of RNA-Seq library to the *S. mansoni* lncRNAs set. TACO assembler was run separately for the RNA-Seq data from samples of four groups: whole organisms (yellow), tissues (red), cell-populations (blue) and single-cells (green), and each value indicates the number of transcripts that were reconstructed specifically with samples from groups indicated in each intersection.
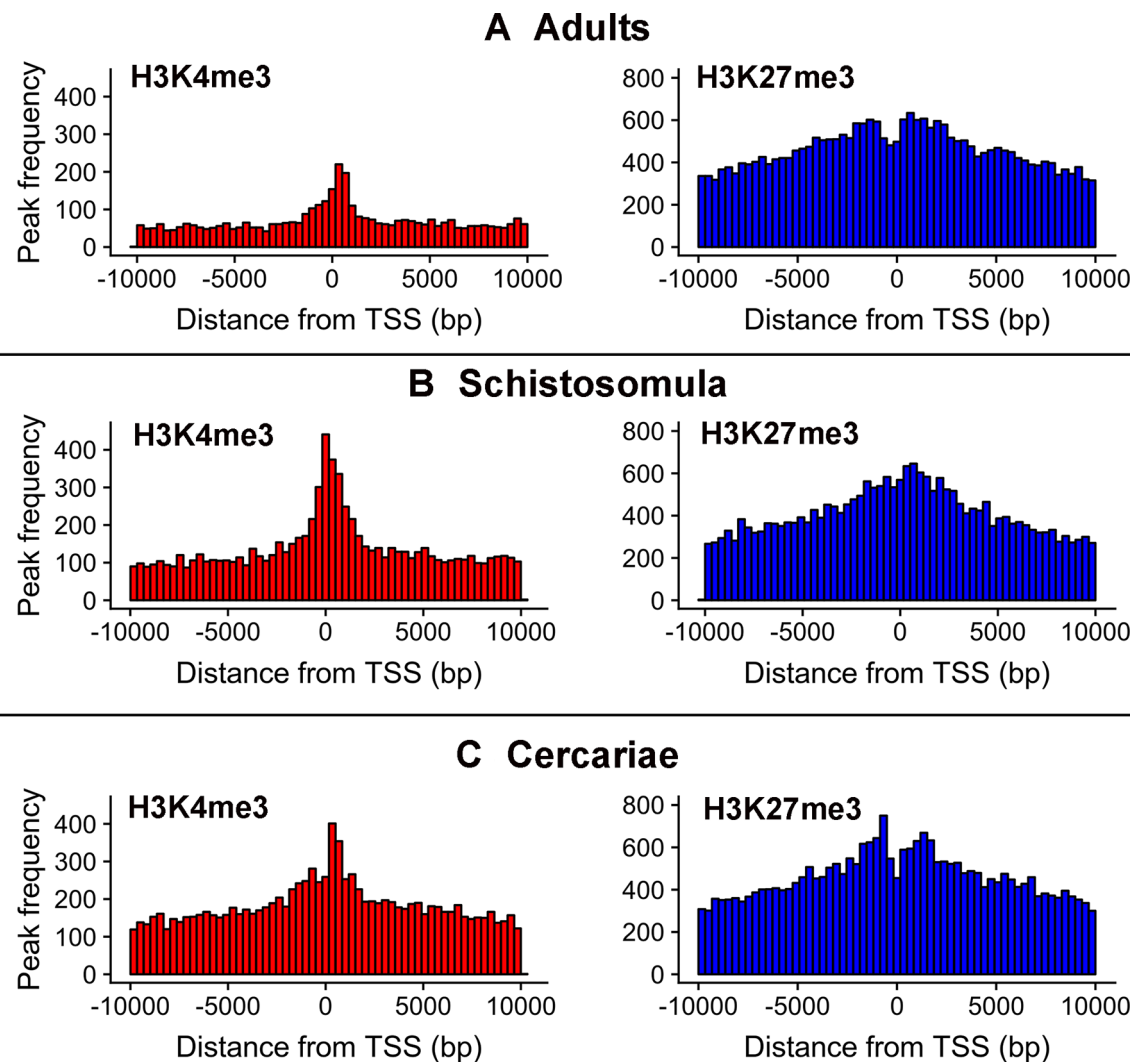
FIGURE 3 | Size distribution in *S. mansoni* of the longest canonical ORFs in lncRNAs and in mRNAs. The graph shows the density (y-axis) of the different sizes for the longest detected ORFs (in nucleotides, x-axis) of all lncRNAs (pink), of all mRNAs (blue) and of their reverse-complement sequences as controls (green and purple, respectively).

similar to that observed by Sati et al. (2012) when comparing histone marks around the TSS of human protein-coding genes and lncRNAs. A total of 8,599 lncRNA transcripts have at least one histone modification mark within 1 kb from their TSS (**Supplementary Table S5**), being 3,659 lincRNAs, 4,188 antisense lncRNAs, and 752 sense lncRNAs. A comparison of the lists of lncRNAs having a given histone mark at their TSS at either of the three different life-cycle stages (**Figure 5**) shows that the most abundant mark is the transcriptional repressive mark, H3K27me3. This mark is present at the TSS of different sets of lncRNAs at each of the three stages, with abundancies ranging from 1,334 lncRNAs with the H3K27me3 mark exclusively in schistosomula to 1,147 lncRNAs with the mark exclusively in adults and 1,024 lncRNAs with the mark exclusively in cercariae (**Figure 5, red**). In addition, the transcriptional activating mark H3K4me3 is present at the
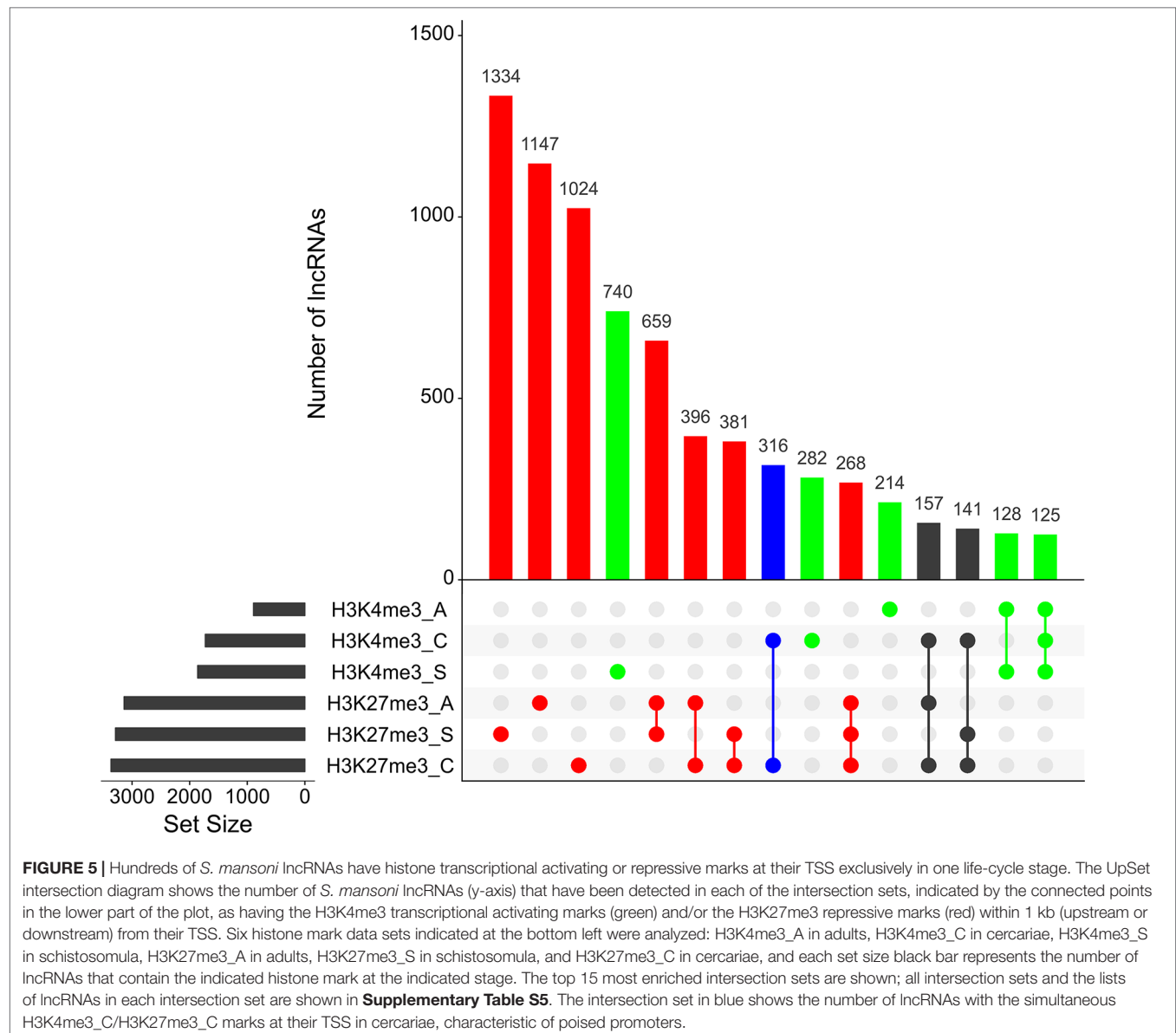
FIGURE 4 | Epigenetic histone marks H3K4me3 and H3K27me3 surrounding the TSS of *S. mansoni* lncRNAs. The frequency of the H3K4me3 marks (red) or of the H3K27me3 marks (blue) mapping within 10 kb around the TSS of all lncRNAs in **(A)** adults, **(B)** schistosomula and **(C)** cercariae was computed.

TSS of a different set of lncRNAs, with abundancies ranging from 740 lncRNAs with the H3K4me3 mark exclusively in schistosomula to 282 lncRNAs with the mark exclusively in cercariae, and 214 lncRNAs with the mark exclusively in adults (**Figure 5, green**). Interestingly, among the lncRNAs with the most abundant patterns of marks at their TSS, there are 316 lncRNAs in cercariae that have the characteristic marks of bivalent poised promoters (having both H3K4me3 and H3K27me3 marks at their TSS) (Voigt et al., 2013) (**Figure 5, blue**). This is analogous to the marks at the TSS of protein-coding genes in cercariae, where most genes have the bivalent mark (Roquis et al., 2018), indicating that lncRNAs are under a similar transcriptional regulatory program as the protein-coding genes in cercariae. **Supplementary Table S5** has a complete UpSet plot similar to that of **Figure 5**, showing the number of lncRNAs found in all different intersections, along with the lists of lncRNAs belonging to each intersection set.

Gene Co-Expression Analyses

Once we identified our final lncRNAs set, we applied weighted gene co-expression network analyses (WGCNA) to integrate the expression level differences observed for lncRNAs and mRNAs among all life-cycle stages and the gonads, using all RNA-seq libraries available. The file containing expression levels (in TPM) for all transcripts in all 633 RNA-Seq libraries is available at http://schistosoma.usp.br/. After normalization and gene filtering (see Methods), 90 libraries out of the 112 from the different stages (mixed-sex adults were not included) remained in the WGCNA analyses, and 19,258 transcripts were retained (12,693 protein-coding genes and 6,565 lncRNAs).

Samples from miracidia, sporocysts, schistosomula, cercariae, and gonads (testes and ovaries) were correctly clustered together by their expression correlation, based on Euclidian distance metrics (**Figure 6**). For samples from adult worms, in spite of the fact that we have one cluster branch mainly composed of

FIGURE 5 | Hundreds of *S. mansoni* lncRNAs have histone transcriptional activating or repressive marks at their TSS exclusively in one life-cycle stage. The UpSet intersection diagram shows the number of *S. mansoni* lncRNAs (y-axis) that have been detected in each of the intersection sets, indicated by the connected points in the lower part of the plot, as having the H3K4me3 transcriptional activating marks (green) and/or the H3K27me3 repressive marks (red) within 1 kb (upstream or downstream) from their TSS. Six histone mark data sets indicated at the bottom left were analyzed: H3K4me3_A in adults, H3K4me3_C in cercariae, H3K4me3_S in schistosomula, H3K27me3_A in adults, H3K27me3_S in schistosomula, and H3K27me3_C in cercariae, and each set size black bar represents the number of lncRNAs that contain the indicated histone mark at the indicated stage. The top 15 most enriched intersection sets are shown; all intersection sets and the lists of lncRNAs in each intersection set are shown in **Supplementary Table S5**. The intersection set in blue shows the number of lncRNAs with the simultaneous H3K4me3_C/H3K27me3_C marks at their TSS in cercariae, characteristic of poised promoters.

females, and another mainly composed of males, there are some male samples among the female ones, and vice versa. Besides, due to the known similarity between males and juveniles (Wang et al., 2017), their samples were not well separated. It is interesting to note that immature females, which were shown to have a similar expression profile as that of males (Lu et al., 2016), are clustered here in the male branch. As the WGCNA performs an unsupervised co-expression analysis, we decided to keep all male and female samples in the analysis, including those that are clustered apart from their main group, in order not to add a bias in the construction of modules.

We identified 15 different lncRNAs/mRNAs co-expression modules **(Figure 7)**, the sizes ranging from 215 to 3,318 transcripts **(Table 2** and **Supplementary Table S6)**. The ratio between the number of lncRNAs and mRNAs that comprise

each module varies among the modules; thus, whereas lncRNAs comprise 86% of the transcripts in the cyan module, only 5% of the transcripts from the black module are lncRNAs **(Table 2)**.
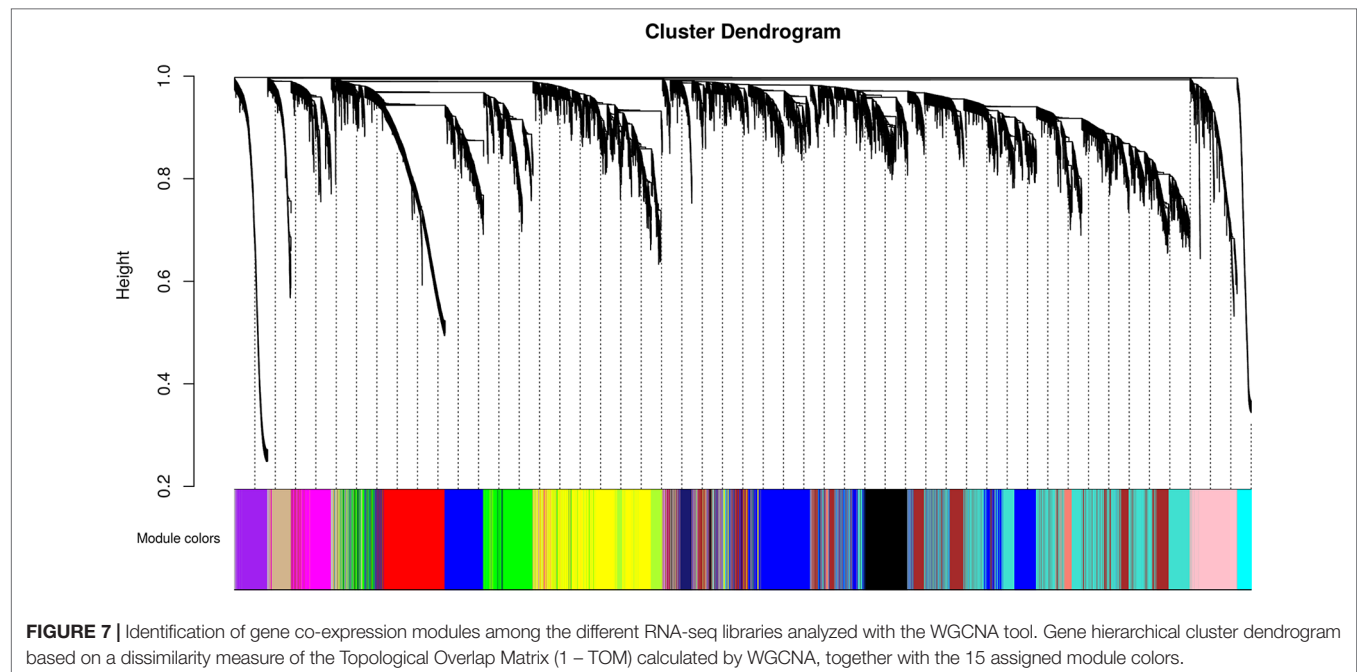
A Pearson correlation analysis indicates that each stage/tissue has at least one module whose gene expression has a statistically significant positive correlation with that stage or tissue **(Figure 8)**. Some stages also have modules that have a statistically significant negative correlation, such as the black module that is negatively correlated with miracidia/sporocysts. For the black module, the transcripts that compose the module have an expression in miracidia/sporocysts that is lower when compared with the overall expression of those transcripts across the other stages. The gray color represents the group of transcripts with a highly heterogeneous co-expression pattern that could not cluster into any of the 15 modules. In fact, it can be seen in **Figure 8** that in

FIGURE 6 | RNA-seq samples clustering based on Euclidian distance according to the expression levels of all genes used for WGCNA. The expression levels of all lncRNA and mRNA genes from all RNA-seq data sets analyzed in the WGCNA were used for unsupervised clustering of the samples, including RNA-seq data sets from adult females (pink), adult males (turquoise), cercariae (tan), immature females (gold), juveniles (midnight blue), miracidia (purple), ovaries (cyan), schistosomula (magenta), sporocysts (green), and testes (red). The SRA or ENA accession number for each RNA-seq library is indicated at each leaf. The asterisks mark the adult male data sets, whose clustering pattern is the most spread one.

this group, the best correlation coefficient obtained in juveniles is lower ($|r| = 0.32$), and the p-value is much higher ($p = 0.002$) than the best parameters that were obtained in at least one stage for any module ($|r| \geq 0.51$ and $p \leq$ 3e-07). Here, our choice of keeping in the WGCNA analysis, those male and female samples that cluster apart from their main group (**Figure 6**) have an impact, decreasing the correlation coefficient of the modules mostly correlated to males or females (pink or turquoise, respectively) when compared with correlation coefficients in the other stages/tissues, nevertheless, they still have a statistically significant high correlation.

We chose three RNA-seq library samples from each of the nine different stages/tissues (among all the libraries under

analysis) to construct a representative expression heatmap (**Figure 9**). This heatmap shows the expression across all stages of the top 50 transcripts with the highest gene module membership (GMM) to the most correlated module of each stage (as seen in **Figure 8)** (for GMM definition see WGCNA background and glossary, available at https://horvath.genetics. ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/ Tutorials/) (Langfelder and Horvath, 2008). The heatmap (**Figure 9**) confirms that the top transcripts belonging to one module are more expressed in one given stage/tissue, which is the stage/tissue with which the module has the highest correlation. It is noteworthy that female library SRR5170160, which clustered inside the male group (**Figure 6**) when all

FIGURE 7 | Identification of gene co-expression modules among the different RNA-seq libraries analyzed with the WGCNA tool. Gene hierarchical cluster dendrogram based on a dissimilarity measure of the Topological Overlap Matrix (1 – TOM) calculated by WGCNA, together with the 15 assigned module colors.

TABLE 2 | Number of transcripts per module and percentage of lncRNAs in each module.

Module color	Total number of transcripts	mRNAs	lncRNAs	% of lncRNAs	Stage of higher absolute correlation value
Black	989	940	49	5	Miracidia/Sporocysts
Blue	3,211	2,688	523	16	Juveniles
Brown	2,466	1,761	705	29	Gonads
Cyan	253	36	217	86	Ovaries
Green	1,308	841	467	35	Gonads
Greenyellow	502	417	85	17	Gonads
Magenta	748	273	475	64	Schistosomula
Midnight blue	215	43	172	80	Juveniles
Pink	840	506	334	40	Adult Females
Purple	590	274	316	54	Miracidia/Sporocysts
Red	1,254	333	921	73	Testes
Salmon	267	230	37	14	Gonads
Tan	356	158	198	56	Cercariae
Turquoise	3,318	2,470	848	26	Adult Males
Yellow	2,067	1,398	669	32	Adult Males

filtered transcripts under analysis were used for clustering, now is correctly clustered with the other female samples (**Figure 9**) when only the top 50 transcripts with the highest GMM are considered. Also, juveniles share with adult males a similar expression pattern of the top 50 male genes, which is in line with the clustering of juveniles along with males in the analysis of **Figure 6**.
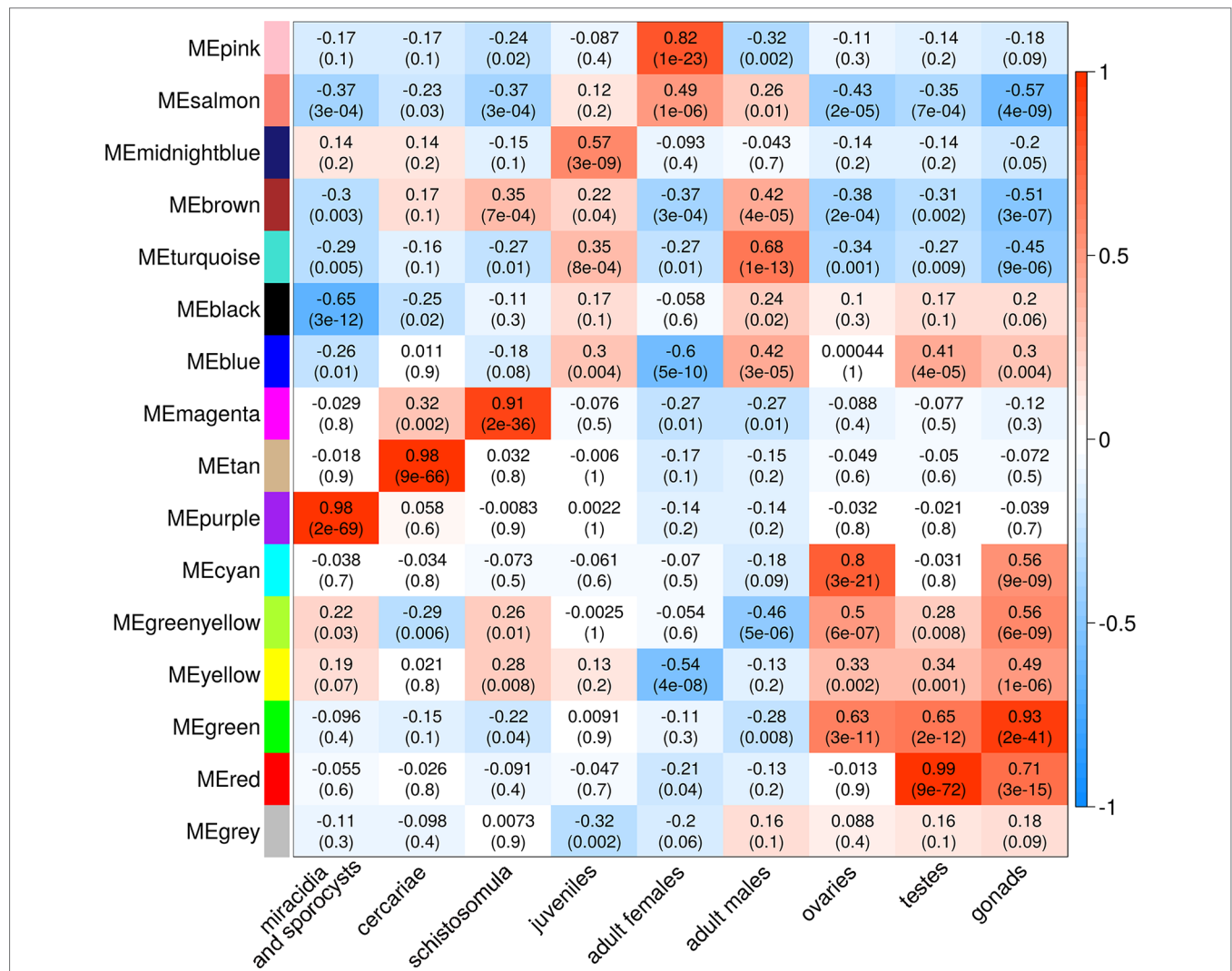
Validation of lncRNAs Expression by RT-qPCR

We designed PCR primer pairs for a selected set of eleven lincRNAs belonging to five different modules, as determined by WGCNA, to detect their expression along the different *S. mansoni* life-cycle stages and to eventually validate their different expression levels at the stages. Our selection was based on the Gene Trait Significance score (GS score) (**Supplementary Table S7**) of each lncRNA in the module where it belongs, which varies from −1 to 1, using the stages as external information (see Methods). The higher the absolute value of the GS score, the more biologically significant and correlated to the stage of interest is the transcript expression. For the RT-qPCR assays, we used samples from eggs (E), miracidia (Mi), cercariae (C), schistosomula (S), adult males (M), and females (F).

First, we measured the expression of five protein-coding genes that were used as stage markers (Parker-Manuel et al.,

FIGURE 8 | Each parasite life-cycle stage (or tissue) has at least one highly correlated gene co-expression module. Each cell in the table shows the Pearson correlation (with the p-value in parenthesis) between each of the 15 co-expression modules determined by WGCNA (indicated at left) and the stages/tissues of *S. mansoni* (indicated at the bottom). The cells are colored according to the scale (at right), which is related to the Pearson correlation coefficient.
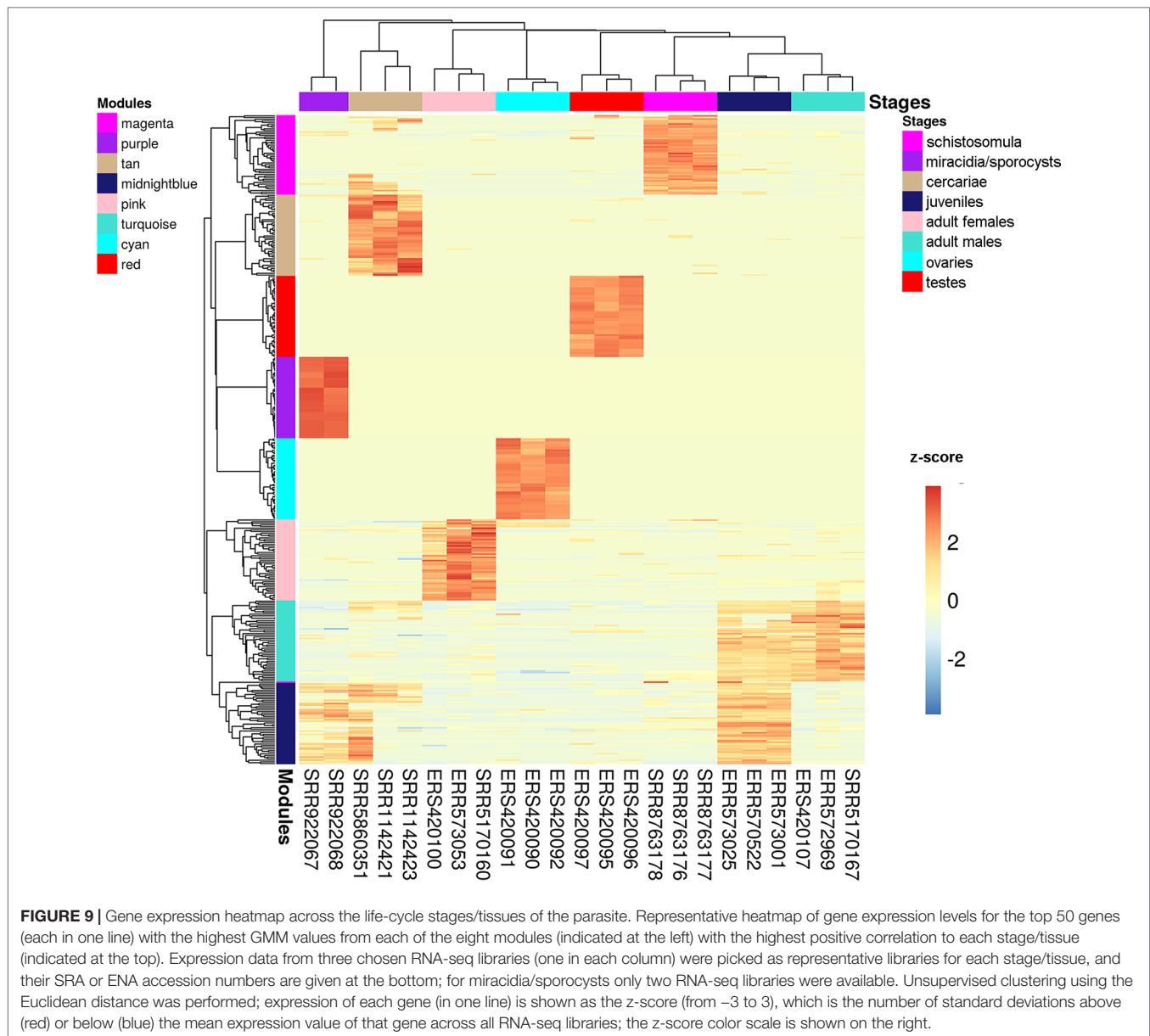
2011; Anderson et al., 2016), and we found that in our RNA samples, they were more highly expressed at the predicted stages (**Supplementary Figure S4**).

Then, we tested the selected eleven lincRNAs and detected that they were expressed in at least one of the six stages that were assayed; specifically, each of six lincRNAs were more highly expressed at the stage predicted by the correlation with the modules (**Figure 10**), at four life-cycle stages: two more highly expressed in miracidia (SmLINC158013-IBu and SmLINC123205-IBu, purple module), two in cercariae (SmLINC123474-IBu and SmLINC134196-IBu, tan module), one in schistosomula (SmLINC105065-IBu, magenta module), and one in males (SmLINC100046-IBu, turquoise module) (**Figure 10**). In **Supplementary Figure S5** we present the values in transcripts per million reads (TPM) from the RNA-seq libraries for each of these six validated lincRNAs. Additionally,

the five other lincRNAs that were tested were detected as expressed across all stages; however, they were not differentially expressed as predicted by the RNA-seq (**Supplementary Figure S6**). This indicates that there is variability of lncRNAs expression between the experimental conditions and parasite strain used in our assays and those found among the dozens of samples that are publicly available.

Protein-Coding Genes Ontology Enrichment and lncRNA Hub Genes in the Modules
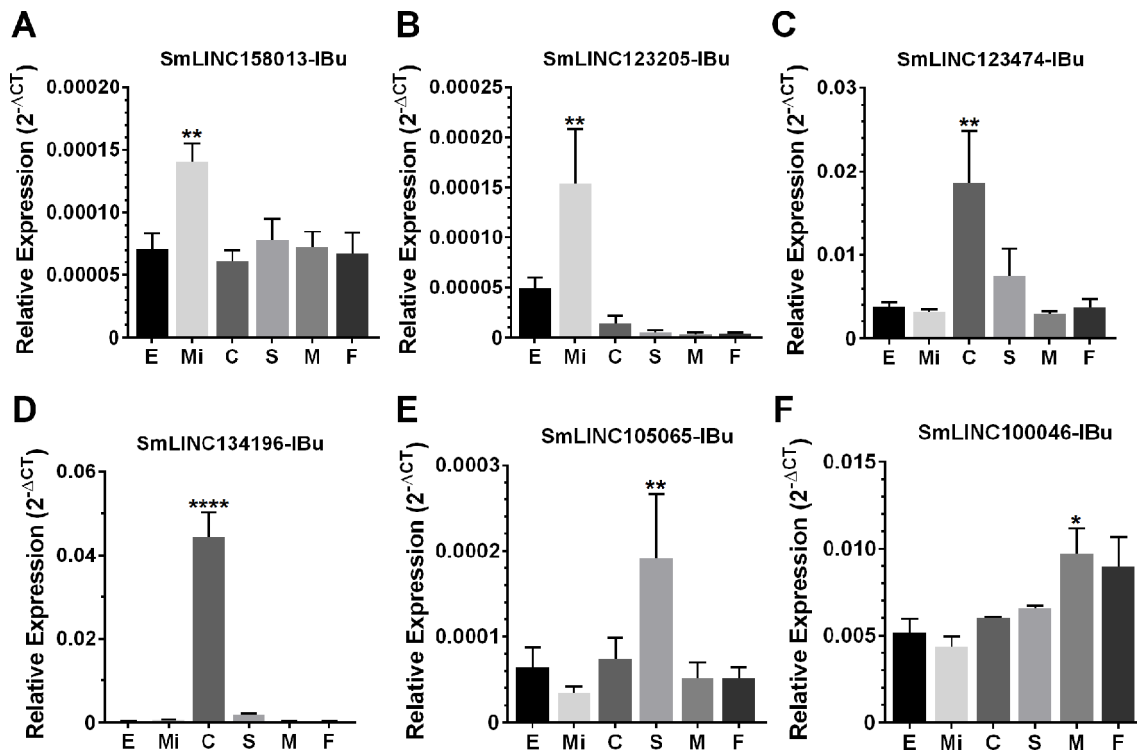
Gene ontology (GO) enrichment analyses show that the protein-coding genes belonging to the red module, which have a correlation of 0.99 with testes, are enriched with processes related to sperm motility such as cilium movement and the

FIGURE 9 | Gene expression heatmap across the life-cycle stages/tissues of the parasite. Representative heatmap of gene expression levels for the top 50 genes (each in one line) with the highest GMM values from each of the eight modules (indicated at the left) with the highest positive correlation to each stage/tissue (indicated at the top). Expression data from three chosen RNA-seq libraries (one in each column) were picked as representative libraries for each stage/tissue, and their SRA or ENA accession numbers are given at the bottom; for miracidia/sporocysts only two RNA-seq libraries were available. Unsupervised clustering using the Euclidean distance was performed; expression of each gene (in one line) is shown as the z-score (from −3 to 3), which is the number of standard deviations above (red) or below (blue) the mean expression value of that gene across all RNA-seq libraries; the z-score color scale is shown on the right.

axoneme assembly **(Figure 11A)**. Besides, the green module, correlated with both ovaries and testes, is enriched with proteins associated with cellular replication **(Figure 11B)**. All other modules with GO enrichment, which in general are enriched with proteins associated to general metabolism, are presented in **Supplementary Figures S7–S10**. The black, cyan, midnight blue, purple, and tan modules have no significantly enriched GO terms due to the small number of protein-coding genes with GO annotation within each of these modules.

All transcripts that belong to the same module are connected; however, to better visualize this, gene co-expression networks were constructed only with the most connected genes (as determined by the adjacency threshold) **(Figure 12)**, and they show, along with the correlation values presented in **Supplementary Table S7**, that some lncRNAs are hub genes

from the network. **Figures 12A**, **B** show lncRNA hub genes in the co-expression networks from the purple and tan modules strongly correlated with miracidia/sporocysts and cercariae life-cycle stages, respectively. In both modules, the lncRNAs represent around half of the transcripts that comprise the modules (see **Table 2**). However, there are some cases, such as in the red module, where three quarters of the member transcripts are lncRNAs, and among the most connected genes in that co-expression network, almost all are lncRNAs **(Figure 12C)**. Also, in the blue module only, 16% of the member transcripts are lncRNAs, and only one is among the most connected genes in the co-expression network **(Figure 12D)**. All the gene networks for all modules in a format compatible with Cytoscape are available at **Supplementary Table S8**. An adjacency cutoff threshold of 0.1 was used.

FIGURE 10 | Confirmation by RT-qPCR of the module-specific lincRNAs relative expression. Six lincRNAs were measured at different developmental stages of *S. mansoni*. From left to right in the x-axis, lincRNAs were measured in RNA samples from eggs (E), miracidia (Mi), cercariae (C), *in vitro* mechanically transformed schistosomula cultivated for 24 h (S), adult males (M) and females (F). The lincRNAs relative gene expression was calculated against the geometric mean of two housekeeping genes: Smp_090920 and Smp_062630. **(A)** and **(B)** show SmLINC158013-IBu and SmLINC123205-IBu representing the **purple** module, specific for miracidia/sporocysts. In **(C)** and **(D),** the SmLINC123474-IBu and SmLINC134196-IBu representing the cercariae-specific **tan** module. In **(E),** the schistosomula-specific lincRNA SmLINC105065-IBu from the **magenta** module and **(F)** the adult male-specific lincRNA SmLINC100046-IBu from the **turquoise** module. Bars represent standard deviation of the mean from four biological replicates for each stage. Two technical replicates were assayed for each of the four biological replicates per stage. The ANOVA Tukey test was used to calculate the statistical significance of the expression differences among the parasite stage samples (*p value≤0.05; **p value≤0.01; ****p value≤0.0001). For clarity purposes, we show only the highest p value obtained in the ANOVA Tukey test for expression comparisons against one another among the stages.

LncRNAs Expressed in Single Cells
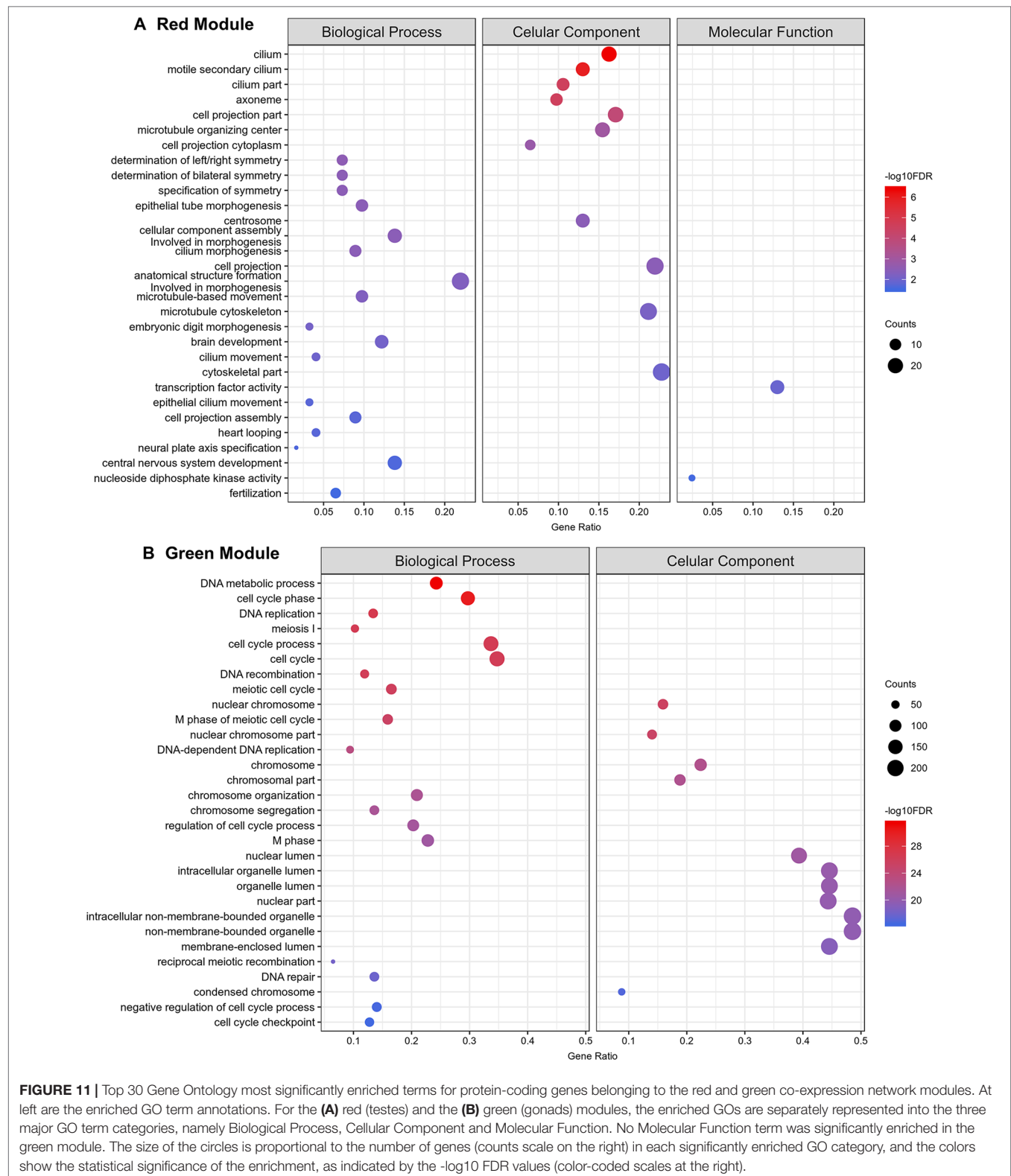
Finally, analyses using single-cell data from two stages, mother sporocysts stem cells and juveniles' stem cells, identified three different clusters. Cluster 1 is composed of a subgroup of juvenile stem cells, cluster 2 is composed of all mother sporocysts stem cells, and cluster 3 is composed of a second and smaller subgroup of juvenile stem cells (**Figure 13A**). The marker gene analyses show, for the first time in *S. mansoni,* that lncRNAs have specific expression also at the single-cell level, where from the top 10 markers that allow us to differentiate mother sporocysts stem cells from juvenile stem cells, eight are lncRNAs (**Figure 13B**), confirming the stage specificity of lncRNAs also seen in whole worm analyses by WGCNA. Besides, another lncRNA was identified as a marker for cluster 3 when compared with the other two clusters (**Figure 13B**).

DISCUSSION

When the human genome was first sequenced, the vast genomic regions that lie between protein-coding genes (intergenic regions) were considered junk DNA; one decade later, the Encyclopedia of DNA Elements (ENCODE) project found that 80% of the human genome serves some biochemical purpose (Pennisi, 2012), including giving rise to the transcription of nearly 10,000 lncRNAs (Derrien et al., 2012). Although we are still at the beginning of the studies with lncRNAs, with the vast majority of their roles and mechanisms of action in human beings still unknown, it is now clear that most of the lncRNAs are transcribed from intergenic regions and are key regulators in vital processes (Kitagawa et al., 2013; Rosa and Ballarino, 2016; Golicz et al., 2018), being associated to several pathologies in humans, such as cancer (Fang and Fullwood, 2016), Alzheimer's (Zijian, 2016), and cardiac diseases (Simona et al., 2018).

In *S. mansoni*, with the release in 2012 of version 5.2 of the genome and annotated transcriptome (Protasio et al., 2012), and with the accumulation until 2017 of large amounts of information on gene expression obtained through 88 publicly available RNA-seq libraries, our group decided to map the RNA-seq data and identify the lncRNAs repertoire expressed in this parasite (Vasconcelos et al., 2017); this was followed by two other papers that provided an additional set of lncRNAs

FIGURE 11 | Top 30 Gene Ontology most significantly enriched terms for protein-coding genes belonging to the red and green co-expression network modules. At left are the enriched GO term annotations. For the **(A)** red (testes) and the **(B)** green (gonads) modules, the enriched GOs are separately represented into the three major GO term categories, namely Biological Process, Cellular Component and Molecular Function. No Molecular Function term was significantly enriched in the green module. The size of the circles is proportional to the number of genes (counts scale on the right) in each significantly enriched GO category, and the colors show the statistical significance of the enrichment, as indicated by the -log10 FDR values (color-coded scales at the right).

(Liao et al., 2018; Oliveira et al., 2018). In the present work, by extending the analysis to 633 publicly available RNA-seq libraries, and by performing a detailed curation of the assembled transcripts, we observed that at the sequencing depth obtained with the current RNA-seq data sets, a considerable amount of partially processed pre-mRNAs is being sequenced. These pre-mRNAs give rise to assembled transcript units showing intron retention and frequent stop codons in the retained introns,

FIGURE 12 | Gene co-expression networks. The top most highly connected genes from the **(A)** purple (miracidia/sporocysts), **(B)** tan (cercariae), **(C)** red (testes) or **(D)** blue (females) co-expression network modules are shown. The adjacency thresholds are 0.77, 0.48, 0.58, and 0.26, respectively. Each red circle represents one lncRNA (SmLINC, SmLNCA, or SmLNCS), and each blue circle represents one protein-coding gene (Smp_). Circle sizes are related to the intramodular connectivity value for each transcript.

and therefore, these transcripts can be mistakenly annotated as lncRNAs.

In fact, the failure to identify partially processed pre-mRNA in previous publications (Vasconcelos et al., 2017; Liao et al., 2018) may explain the report of probable protein-coding genes as lncRNAs (**Supplementary Figure S1**). Our current pipeline has removed at step 4 a total of 31,183 assembled transcripts that had partial or total exon-exon overlap on the same genomic strand with known *S. mansoni* protein-coding genes, and this included around 14,000 assembled transcripts that represented fully processed mature protein-coding transcripts that exactly matched the annotated v 7.1 transcripts from the Wellcome Sanger Institute, as well as some 17,000 assembled transcripts that for the most part represent partially processed pre-mRNAs with intron retention; among the latter are 4,293 transcripts that

were previously classified as lncRNAs (Vasconcelos et al., 2017; Liao et al., 2018; Oliveira et al., 2018) and are now excluded. With the six stringent filtering steps used in the present work, we are confident that our final set of 16,583 lncRNAs is a robust representation of the lncRNAs complement expressed in *S. mansoni*, of which 11,022 transcripts are novel lncRNAs, and 5,561 have gene overlap with lncRNAs already reported in previous works (Vasconcelos et al., 2017; Liao et al., 2018; Oliveira et al., 2018).

One question that has been raised about lncRNAs is the possibility that their function is executed through translation into short peptides, a concern that arises from the fact that almost all lncRNAs encode short canonical ORFs within their sequences (Verheggen et al., 2017); the fact that the size distribution of ORFs found within our set of lncRNAs (sense) is very similar to

FIGURE 13 | Single-cell expression analysis identified three different cell population clusters when comparing *S. mansoni* juveniles' stem cells and mother sporocysts stem cells and lncRNAs as gene markers at the single-cell level. **(A)** Single-cell RNA-Seq data from two RNA-Seq libraries, one from juveniles' stem cells and another from mother sporocysts' stem cells, were analyzed with the SC3 tool that performed an unsupervised clustering of the cells based on the single-cell gene expression data. Principal component analysis plot, where the symbol colors and sizes indicate the three clusters identified by SC3, and the shapes indicate the two life-cycle stages from which the stem cells were isolated. The symbol size is inversely related to the number of cells that belong to the cluster. **(B)** In the marker-gene expression matrix (log-transformation, represented by the color scale), the statistically significant gene markers are the rows, and the cells are columns. The life-cycle stage from which each cell was isolated is indicated by the color bar at the top (stages). The clusters of cells are separated by white vertical lines and are indicated by the second color bar at the top (clusters). The cluster marker genes are separated by white horizontal lines, the markers groups are indicated at left, and the names of the marker genes at right. Only the top 10 most significant marker genes are shown for cluster 2.

the size distribution of random ORFs found within their reverse-complement sequences and within the reverse-complement sequence of mRNAs suggests that the putative short ORFs from the lncRNAs identified here are indeed random ORFs, most probably not translated into short functional peptides. Nevertheless, future functional characterization in *S. mansoni* of selected lncRNAs may eventually include a search for a possible

dual function role (Nam et al., 2016; Choi et al., 2018) both as lncRNA and through a translated short peptide.

Histone marks were found here at the TSS of lncRNAs, and the identification of different sets of lncRNAs that have at their TSS the transcriptional activation H3K4me3 mark, or the repressive H3K27me3 mark, when the three life-cycle stages are compared, suggests that lncRNAs expression in *S. mansoni*

is regulated by an epigenetic program. This finding reinforces the hypothesis that different lncRNAs may play important roles along the parasite life-cycle, and the sets of lncRNAs identified in this analysis might be the first candidates to be explored for further functional characterization.

Gene co-expression networks correlated to the different *S. mansoni* life-cycle stages were identified by our analyses, and they pointed to sets of protein-coding genes and lncRNAs with expression most correlated to one given stage. This information provides an initial platform for prioritizing the lncRNAs to be selected for further direct functional characterization, which will include a search for altered *S. mansoni* phenotypes upon knockdown of lncRNA candidates. In *Plasmodium falciparum*, the knockdown of antisense lncRNAs has down-regulated the active var gene, a gene related to immune evasion, erasing the epigenetic memory and substantially changing the var gene expression pattern (Amit-Avraham et al., 2015). In analogy, it is expected that characterization of lncRNAs in *S. mansoni* will help to recognize the biochemical pathways where they play a functional role, will permit to identify their interacting protein partners, and will eventually point to relevant ways of intervention in the parasite physiology.

Due to the complex and diverse mechanisms displayed by lncRNAs in regulating protein-coding genes and miRNAs, the majority of studies have not progressed beyond cell or animal models, and progression toward the clinic has been slow (Harries, 2019). Nevertheless, lncRNAs represent potentially good therapeutic targets (Matsui and Corey, 2017; Blokhin et al., 2018; Harries, 2019). As reviewed by Matsui and Corey (2017), in Angelman syndrome model mouse, the administration of antisense oligonucleotides (ASOs), which target the Ube3a-ATS lncRNA for degradation, partially reversed some cognitive defects associated with the disease in the animals (Meng et al., 2014). Also, in xenograft melanoma models, the intravenous injection of ASOs targeting the lncRNA SAMMSON caused p53 activation, tumor growth suppression, decreased cell proliferation, and increased apoptosis (Leucci et al., 2016). In this respect, it is noteworthy that lncRNAs are considerably less conserved between species when compared with protein-coding genes (Pang et al., 2006; Blokhin et al., 2018), and that only a few dozen ancient lncRNAs have conserved orthologs between ancient non-amniote *Xenopus* and the closest amniote chicken model animals (Necsulea et al., 2014), which shows that lncRNAs have evolutionarily conserved gene regulatory functions but low-sequence conservation across distant species (Necsulea et al., 2014). This feature reduces the chances that targeting a lncRNA in *S. mansoni*, for example, with ASOs, will cause unwanted off-target effects against the mammalian host.

DATA AVAILABILITY

The data sets analyzed in this study can be found in the SRA repository (https://www.ncbi.nlm.nih.gov/sra) and in the ENA repository (https://www.ebi.ac.uk/ena). The specific accession numbers for each and all data sets that were downloaded from these databases and used here are given in **Supplementary Table S1**.

ETHICS STATEMENT

AUTHOR CONTRIBUTIONS

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00823/full#supplementary-material

REFERENCES

Amit-Avraham, I., Pozner, G., Eshar, S., Fastman, Y., Kolevzon, N., Yavin, E., et al. (2015). Antisense long noncoding RNAs regulate var gene activation in the malaria parasite *Plasmodium falciparum*. Proc. Nat.l Acad. Sci. U. S. A. 112, E982–E991. doi: 10.1073/pnas.1420855112

Anderson, L., Amaral, M. S., Beckedorff, F., Silva, L. F., Dazzani, B., Oliveira, K. C., et al. (2016). *Schistosoma mansoni* egg, adult male and female comparative gene expression analysis and identification of novel genes by RNA-Seq. *PLoS Negl. Trop. Dis.* 9, e0004334. doi: 10.1371/journal.pntd.0004334

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D. E., Wang, Z., et al. (2007). High-resolution profiling of histone methylations in the human genome. *Cell* 129, 823–837. doi: 10.1016/j.cell.2007.05.009

Basch, P. F. (1976). Intermediate host specificity in *Schistosoma mansoni*. *Exp. Parasitol.* 39, 150–169. doi: 10.1016/0014-4894(76)90022-9

Basch, P. F. (1981). Cultivation of *Schistosoma mansoni in vitro*. I. Establishment of cultures from cercariae and development until pairing. *J. Parasitol.* 67, 179–185. doi: 10.2307/3280632

Batugedara, G., Lu, X. M., Bunnik, E. M., and Le Roch, K. G. (2017). The role of chromatin structure in gene regulation of the human malaria parasite. *Trends Parasitol* 33, 364–377. doi: 10.1016/j.pt.2016.12.004

Bhat, S. A., Ahmad, S. M., Mumtaz, P. T., Malik, A. A., Dar, M. A., Urwat, U., et al. (2016). Long non-coding RNAs: mechanism of action and functional utility. *Non-coding RNA Research* 1, 43–50. doi: 10.1016/j.ncrna.2016.11.002

Blokhin, I., Khorkova, O., Hsiao, J., and Wahlestedt, C. (2018). Developments in lncRNA drug discovery: where are we heading? *Expert Opin. Drug Discov.* 13, 837–849. doi: 10.1080/17460441.2018.1501024

Cao, H., Wahlestedt, C., and Kapranov, P. (2018). Strategies to annotate and characterize long noncoding rnas: advantages and pitfalls. *Trends Genet.* 34, 704–721. doi: 10.1016/j.tig.2018.06.002

Cdc. (2018). *Centers for Disease Control and Prevention, Parasites - Schistosomiasis* [Online]. Available: https://www.cdc.gov/parasites/schistosomiasis/ [Accessed 20/07/2018 2018].

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560

Choi, S. W., Kim, H. W., and Nam, J. W. (2018). The small peptide world in long noncoding RNAs. *Brief. Bioinformatics*, bby055. doi: 10.1093/bib/bby055

Credendino, S. C., Lewin, N., De Oliveira, M., Basu, S., Andrea, B., Amendola, E., et al. (2017). Tissue- and cell type-specific expression of the long noncoding RNA Klhl14-AS in Mouse. *Int. J. Genomics* 2017, 7. doi: 10.1155/2017/9769171

Dalton, J. P., Day, S. R., Drew, A. C., and Brindley, P. J. (1997). A method for the isolation of schistosome eggs and miracidia free of contaminating host tissues. *Parasitology* 115 (Pt 1), 29–32. doi: 10.1017/S0031182097001091

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* 22, 1775–1789. doi: 10.1101/gr.132159.111

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635

Fang, Y., and Fullwood, M. J. (2016). Roles, functions, and mechanisms of long non-coding RNAs in cancer. *Genomics Proteomics Bioinformatics* 14, 42–54. doi: 10.1016/j.gpb.2015.09.006

Golicz, A. A., Bhalla, P. L., and Singh, M. B. (2018). lncRNAs in plant and animal sexual reproduction. *Trends Plant Sci.* 23, 195–205. doi: 10.1016/j.tplants.2017.12.009

Gomes Casavechia, M. T., De Melo, G.D.a.N., Da Silva Fernandes, A. C. B., De Castro, K. R., Pedroso, R. B., Da Silva Santos, T., et al. (2018). Systematic review and meta-analysis on *Schistosoma mansoni* infection prevalence, and associated risk factors in Brazil. *Parasitology* 145, 1000–1014. doi: 10.1017/S0031182017002268

Hanly, D. J., Esteller, M., and Berdasco, M. (2018). Interplay between long non-coding RNAs and epigenetic machinery: emerging targets in cancer? *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 373, 20170074. doi: 10.1098/rstb.2017.0074

Harries, L. W. (2019). RNA biology provides new therapeutic targets for human disease. *Front. Genet.* 10, 205. doi: 10.3389/fgene.2019.00205

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., et al. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* 38, 576–589. doi: 10.1016/j.molcel.2010.05.004

Howe, K. L., Bolt, B. J., Shafie, M., Kersey, P., and Berriman, M. (2017). WormBase ParaSite—a comprehensive resource for helminth genomics. *Mol. Biochem. Parasitol.* 215, 2–10. doi: 10.1016/j.molbiopara.2016.11.005

Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., Von Mering, C., et al. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* 34, 2115–2122. doi: 10.1093/molbev/msx148

Kiselev, V. Y., Kirschner, K., Schaub, M. T., Andrews, T., Yiu, A., Chandra, T., et al. (2017). SC3: consensus clustering of single-cell RNA-seq data. *Nat. Methods* 14, 483. doi: 10.1038/nmeth.4236

Kitagawa, M., Kitagawa, K., Kotake, Y., Niida, H., and Ohhata, T. (2013). Cell cycle regulation by long non-coding RNAs. *Cell. Mol. Life Sci.* 70, 4785–4794. doi: 10.1007/s00018-013-1423-0

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9, 559. doi: 10.1186/1471-2105-9-559

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Leucci, E., Vendramin, R., Spinazzi, M., Laurette, P., Fiers, M., Wouters, J., et al. (2016). Melanoma addiction to the long non-coding RNA SAMMSON. *Nature* 531, 518. doi: 10.1038/nature17161

Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., and Pfister, H. (2014). UpSet: Visualization of intersecting sets. *IEEE Trans. Vis. Comput. Graph.* 20, 1983–1992. doi: 10.1109/TVCG.2014.2346248

Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12, 323. doi: 10.1186/1471-2105-12-323

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. and Genome Project Data Processing, S. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Liao, Q., Zhang, Y., Zhu, Y., Chen, J., Dong, C., Tao, Y., et al. (2018). Identification of long noncoding RNAs in *Schistosoma mansoni* and *Schistosoma japonicum*. *Exp. Parasitol.* 191, 82–87. doi: 10.1016/j.exppara.2018.07.001

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2–DDCT Method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Lu, Z., Sessler, F., Holroyd, N., Hahnel, S., Quack, T., Berriman, M., et al. (2016). Schistosome sex matters: a deep view into gonad-specific and pairing-dependent transcriptomes reveals a complex gender interplay. *Sci Rep.* 6, 31150. doi: 10.1038/srep31150

Maere, S., Heymans, K., and Kuiper, M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21, 3448–3449. doi: 10.1093/bioinformatics/bti551

Matsui, M., and Corey, D. R. (2017). Non-coding RNAs as drug targets. *Nat. Rev. Drug Discov.* 16, 167–179. doi: 10.1038/nrd.2016.117

Mccarthy, D. J., Campbell, K. R., Wills, Q. F., and Lun, A. T. L. (2017). Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* 33, 1179–1186. doi: 10.1093/bioinformatics/btw777

Meng, L., Ward, A. J., Chun, S., Bennett, C. F., Beaudet, A. L., and Rigo, F. (2014). Towards a therapy for Angelman syndrome by targeting a long non-coding RNA. *Nature* 518, 409. doi: 10.1038/nature13975

Nam, J. W., Choi, S. W., and You, B. H. (2016). Incredible RNA: dual functions of coding and noncoding. *Mol. Cells* 39, 367–374. doi: 10.14348/molcells.2016.0039

Necsulea, A., Soumillon, M., Warnefors, M., Liechti, A., Daish, T., Zeller, U., et al. (2014). The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* 505, 635–640. doi: 10.1038/nature12943

Niknafs, Y. S., Pandian, B., Iyer, H. K., Chinnaiyan, A. M., and Iyer, M. K. (2016). TACO produces robust multisample transcriptome assemblies from RNA-seq. *Nat. Methods* 14, 68. doi: 10.1038/nmeth.4078

Oliveira, K. C., Carvalho, M. L., Maracaja-Coutinho, V., Kitajima, J. P., and Verjovski-Almeida, S. (2011). Non-coding RNAs in schistosomes: an unexplored world. *An. Acad. Bras. Cienc.* 83, 673–694. doi: 10.1590/S0001-37652011000200026

Oliveira, V. F., Mota, E. A., Jannotti-Passos, L. K., Coelho, P. M. Z., Mattos, A. C. A., Couto, F. F. B., et al. (2018). Identification of 170 new long noncoding RNAs in *Schistosoma mansoni*. *Biomed Res. Int.* 2018, 1264697. doi: 10.1155/2018/1264697

Pang, K. C., Frith, M. C., and Mattick, J. S. (2006). Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends Genet.* 22, 1–5. doi: 10.1016/j.tig.2005.10.003

Parker-Manuel, S. J., Ivens, A. C., Dillon, G. P., and Wilson, R. A. (2011). Gene expression patterns in larval *Schistosoma mansoni* associated with infection of the mammalian host. *PLoS Negl. Trop. Dis.* 5, e1274. doi: 10.1371/journal.pntd.0001274

Pennisi, E. (2012). Genomics. ENCODE project writes eulogy for junk DNA. *Science* 337, 1159, 1161. doi: 10.1126/science.337.6099.1159

Protasio, A. V., Tsai, I. J., Babbage, A., Nichol, S., Hunt, M., Aslett, M. A., et al. (2012). A systematically improved high quality genome and transcriptome of the human blood fluke *Schistosoma mansoni*. *PLoS Negl. Trop. Dis.* 6, e1455. doi: 10.1371/journal.pntd.0001455

Roquis, D., Taudt, A., Geyer, K. K., Padalino, G., Hoffmann, K. F., Holroyd, N., et al. (2018). Histone methylation changes are required for life cycle progression in the human parasite *Schistosoma mansoni*. *PLOS Pathog.* 14, e1007066. doi: 10.1371/journal.ppat.1007066

Rosa, A., and Ballarino, M. (2016). Long noncoding RNA regulation of pluripotency. *Stem Cells Int.* 2016, 1797692. doi: 10.1155/2016/1797692

Sati, S., Ghosh, S., Jain, V., Scaria, V., and Sengupta, S. (2012). Genome-wide analysis reveals distinct patterns of epigenetic features in long non-coding RNA loci. *Nucleic Acids Res.* 40, 10018–10031. doi: 10.1093/nar/gks776

Shao, M., and Kingsford, C. (2017). Accurate assembly of transcripts through phase-preserving graph decomposition. *Nat. Biotechnol.* 35, 1167. doi: 10.1038/nbt.4020

Simona, G., Antonio, S. S., Yvan, D., and Fabio, M., (2018). Long noncoding RNAs and cardiac disease. *Antioxid. Redox. Signal.* 29, 880–901. doi: 10.1089/ars.2017.7126

Sims, D., Sudbery, I., Ilott, N. E., Heger, A., and Ponting, C. P. (2014). Sequencing depth and coverage: key considerations in genomic analyses. *Nat. Rev. Genet.* 15, 121–132. doi: 10.1038/nrg3642

Tarashansky, A. J., Xue, Y., Quake, S. R., and Wang, B. (2018). Self-assembling manifolds in single-cell RNA sequencing data. *bioRxiv.* doi: 10.1101/364166

Vasconcelos, E. J. R., Dasilva, L. F., Pires, D. S., Lavezzo, G. M., Pereira, A. S. A., Amaral, M. S., et al. (2017). The *Schistosoma mansoni* genome encodes thousands of long non-coding RNAs predicted to be functional at different parasite life-cycle stages. *Sci Rep.* 7, 10508. doi: 10.1038/s41598-017-10853-6

Verheggen, K., Volders, P.-J., Mestdagh, P., Menschaert, G., Van Damme, P., Gevaert, K., et al. (2017). Noncoding after all: biases in proteomics data do not explain observed absence of lncRNA translation products. *J. Proteome Res.* 16, 2508–2515. doi: 10.1021/acs.jproteome.7b00085

Voigt, P., Tee, W.-W., and Reinberg, D. (2013). A double take on bivalent promoters. *Genes Dev.* 27, 1318–1338. doi: 10.1101/gad.219626.113

Wang, B., Lee, J., Li, P., Saberi, A., Yang, H., Liu, C., et al. (2018). Stem cell heterogeneity drives the parasitic life cycle of *Schistosoma mansoni*. *Elife* 7, e35449. doi: 10.7554/eLife.35449

Wang, J., Yu, Y., Shen, H., Qing, T., Zheng, Y., Li, Q., et al. (2017). Dynamic transcriptomes identify biogenic amines and insect-like hormonal regulation for mediating reproduction in *Schistosoma japonicum*. *Nat. Commun.* 8, 14693. doi: 10.1038/ncomms14693

Wang, L., Wang, S., and Li, W. (2012). RSeQC: quality control of RNA-seq experiments. *Bioinformatics* 28, 2184–2185. doi: 10.1093/bioinformatics/bts356

Who, W. H. O. (2015). *Investing to overcome the global impact of neglected tropical diseases: third WHO report on neglected tropical diseases 2015*. Geneva, Switzerland: World Health Organization.

Wu, W., Wagner, E. K., Hao, Y., Rao, X., Dai, H., Han, J., et al. (2016). Tissue-specific co-expression of long non-coding and coding RNAs associated with breast cancer. *Sci Rep.* 6, 32731. doi: 10.1038/srep32731

Wucher, V., Legeai, F., Hédan, B., Rizk, G., Lagoutte, L., Leeb, T., et al. (2017). FEELnc: a tool for long non-coding RNA annotation and its application to the dog transcriptome. *Nucleic Acids Res.* 45, e57–e57. doi: 10.1093/nar/gkw1306

Yang, D.-C., Kong, L., Wei, L., Hou, M., Kang, Y.-J., Meng, Y.-Q., et al. (2017). CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* 45, W12–W16. doi: 10.1093/nar/gkx428

Zijian, Z. (2016). Long non-coding RNAs in Alzheimer's disease. *Curr. Top. Med. Chem.* 16, 511–519. doi: 10.2174/1568026615666150813142956

Zoni, A. C., Catalá, L., and Ault, S. K. (2016). Schistosomiasis prevalence and intensity of infection in Latin America and the Caribbean Countries, 1942-2014: a systematic review in the context of a regional elimination goal. *PLoS Negl. Trop. Dis.* 10, e0004493. doi: 10.1371/journal.pntd.0004493

Check for
updates

Evaluation of DNA Extraction Methods on Individual Helminth Egg and Larval Stages for Whole-Genome Sequencing

Stephen R. Doyle[1]*, Geetha Sankaranarayanan[1], Fiona Allan[2], Duncan Berger[1], Pablo D. Jimenez Castro[3,4], James Bryant Collins[3], Thomas Crellen[1,5], María A. Duque-Correa[1], Peter Ellis[1], Tegegn G. Jaleta[6], Roz Laing[7], Kirsty Maitland[7], Catherine McCarthy[1], Tchonfienet Moundai[8], Ben Softley[1], Elizabeth Thiele[9], Philippe Tchindebet Ouakou[8], John Vianney Tushabe[1,10], Joanne P. Webster[11], Adam J. Weiss[12], James Lok[6], Eileen Devaney[7], Ray M. Kaplan[3], James A. Cotton[1], Matthew Berriman[1] and Nancy Holroyd[1]*

[1] Parasites and Microbes, Wellcome Sanger Institute, Hinxton, United Kingdom, [2] Department of Life Sciences, Natural History Museum, London, United Kingdom, [3] Department of Infectious Diseases, College of Veterinary Medicine, University of Georgia, Athens, GA, United States, [4] Grupo de Parasitologia Veterinaria, Universidad Nacional de Colombia, Bogotá, Colombia, [5] Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom, [6] Department of Pathobiology, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, PA, United States, [7] Institute of Biodiversity Animal Health and Comparative Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom, [8] Ministry of Public Health, N'Djamena, Chad, [9] Department of Biology, Vassar College, Poughkeepsie, NY, United States, [10] Medical Research Council/Uganda Virus Research Institute and London School of Hygiene & Tropical Medicine Uganda Research Unit, Entebbe, Uganda, [11] Centre for Emerging, Endemic and Exotic Diseases, Department of Pathology and Population Sciences, Royal Veterinary College, University of London, Herts, United Kingdom, [12] Guinea Worm Eradication Program, The Carter Center, Atlanta, GA, United States

Whole-genome sequencing is being rapidly applied to the study of helminth genomes, including *de novo* genome assembly, population genetics, and diagnostic applications. Although late-stage juvenile and adult parasites typically produce sufficient DNA for molecular analyses, these parasitic stages are almost always inaccessible in the live host; immature life stages found in the environment for which samples can be collected non-invasively offer a potential alternative; however, these samples typically yield very low quantities of DNA, can be environmentally resistant, and are susceptible to contamination, often from bacterial or host DNA. Here, we have tested five low-input DNA extraction protocols together with a low-input sequencing library protocol to assess the feasibility of whole-genome sequencing of individual immature helminth samples. These approaches do not use whole-genome amplification, a common but costly approach to increase the yield of low-input samples. We first tested individual parasites from two species spotted onto FTA cards—egg and L1 stages of *Haemonchus contortus* and miracidia of *Schistosoma mansoni*—before further testing on an additional five species—*Ancylostoma caninum*, *Ascaridia dissimilis*, *Dirofilaria immitis*, *Strongyloides stercoralis*, and *Trichuris muris*—with an optimal protocol. A sixth species—*Dracunculus medinensis*—was included for comparison. Whole-genome sequencing followed by analyses to determine the proportion of on- and off-target mapping revealed successful sample preparations for six of the eight species tested with variation both between species and between different life stages from some species described. These results demonstrate the feasibility of whole-genome sequencing of individual parasites, and highlight a new avenue toward

generating sensitive, specific, and information-rich data for the diagnosis and surveillance of helminths.

INTRODUCTION

Accurate methods for diagnosis and surveillance of helminth infections are of increasing interest in both human and animal health settings. Such approaches are typically proposed to monitor the presence and ultimately decline of populations targeted by large-scale control measures, such as mass drug administration (MDA) for the prevention and/or treatment of human helminth infections, or prophylactic treatment of domesticated animals. An ideal diagnostic will be sensitive to detect the parasite if in fact present, and specific, to identify the targeted parasite species in the presence of non-target material, such as other parasite species or the host. Ideally, samples taken for diagnostic purposes could be used to gather additional information beyond the presence or absence of a specific parasite, so the same material could be used, for example, to predict how well the infection will respond to drug treatment or how the parasite is related to other endemic or imported parasites. As most parasitic stages of helminths of humans and animals are naturally inaccessible *in vivo* (not accounting for potential availability of some mature stages of helminths following chemo-expulsion, for example, *Ascaris lumbricoides* and *Trichuris trichiura*), a diagnostic should also be informative on non-invasive stages of the parasite, such as eggs deposited in feces, or intermediate stages of the parasite's life cycle that exist in the external environment.

A key challenge of working with environmental stages of helminth parasites is that they are often immature, for example, eggs or early stage larvae, and extremely small (for example, *Haemonchus contortus* eggs are approximately 75 × 44 µm and *Schistosoma mansoni* miracidia approximately 140 × 55 µm), limiting the amount of accessible material (e.g., DNA) available to be assayed. They are often environmentally resistant, and the same features that naturally protect the DNA from damage prior to reinfection make it difficult to extract DNA. In many cases, they are isolated from host feces and so are susceptible to bacterial contamination or from host tissues and so become contaminated with host DNA. Furthermore, samples may need to be transported efficiently to a laboratory setting without a significant loss of this already limited material. A number of approaches have been tested to preserve macromolecules from individual parasites for transport and storage, including ethanol, RNAlater, and Whatman® FTA® cards, from which robust PCR and microsatellite data could be profiled (Gower et al., 2007; Webster, 2009; Webster et al., 2012; Xiao et al., 2013; Marek et al., 2014; Boué et al., 2017; Campbell et al., 2017). Although, under ideal conditions, the detection of a single DNA molecule is possible, the limited material available per parasite has, to date, largely restricted assaying to a small number of loci, limiting the amount of information obtained from any individual parasite.

Genomic approaches offer an information-rich technology for diagnostic and surveillance applications. Increasing throughput and decreasing costs of whole-genome sequencing has resulted in the recent and steadily growing application of genomics in helminth parasitology, for example, for diagnostic applications, high-throughput amplicon sequencing for helminth species identification and community composition (Avramenko et al., 2015) and the presence of drug resistance alleles (Avramenko et al., 2019) have been described. Although low DNA concentrations are typically prohibitive for genome-wide approaches on individual parasites, a number of studies have successfully used whole-genome amplification on DNA extracted from single larval stages to perform reduced representation (Shortt et al., 2017) and exome (Le Clec'h et al., 2018; Platt et al., 2019) sequencing on miracidia of *Schistosoma* spp., and whole-genome sequencing of *Haemonchus contortus* L3 stage larvae (Doyle et al., 2018) and microfilariae of *Wuchereria bancrofti* (Small et al., 2018). Whole-genome amplification protocols do, however, add considerable expense per sample and can introduce technical artefacts, such as uneven and/or preferential amplification (potentially of contaminant sequences), chimeric sequences, and allele dropout (Tsai et al., 2014; Sabina and Leamon, 2015), that may lead to a reduction in genetic diversity, and in turn, relevance to the original unamplified material. The field of genomics is, however, rapidly advancing toward very low minimum sample input requirements, and single-cell protocols for DNA and RNA sequencing are now available. Such approaches have begun to be used on parasitic species, such as *Plasmodium* spp. (Trevino et al., 2017; Ngara et al., 2018; Reid et al., 2018; Howick et al., 2019), but are yet to be adopted by helminth parasitologists. Although these low-input, high-throughput approaches are not designed—and perhaps not currently suitable—for diagnostic applications, the development of molecular biology techniques for low-input sequencing could aid in the use of genomics for helminth applications. Here, we tested a number of low-input DNA extraction approaches for individual helminth samples stored on Whatman® FTA® cards, followed by low-input library preparation without whole-genome amplification and whole-genome sequencing. A total of five DNA extraction approaches were initially tested, after which the most promising approach was applied to multiple life stages from seven helminth species (with an additional species presented for comparison). The results presented here demonstrate the advancement of low-input whole-genome sequencing and may be broadly applicable to helminth and non-helminth species for which low-input whole-genome sequencing is to be performed. Finally, we discuss our results in the context of helminth diagnostics and surveillance.

METHODS

Sample Collection

Samples representing accessible, immature life stages of a total of eight helminth species were tested, the collection of which is described below. A number of protocols and substrates or solutions have been described for the safe, effective storage of samples for downstream molecular biology applications; we chose Whatman® FTA® cards as a substrate to store and transport parasite material, as they are a relatively cost-effective method for storing samples without the need for specialized storage and transport conditions, for example, a cold chain from collection to sequencing. Removal of the logistical hurdles and their associated costs is particularly important for collection of specimens from endemic regions, for example, *S. mansoni* miracidia that were isolated from patients and purified in Africa before transported and stored on FTA cards for down-stream processing the UK as described here.

Ancylostoma caninum: Fresh feces from a research purpose-bred laboratory beagle (University of Georgia, AUP A2017 10-016-Y1-A0) infected with the Barrow isolate (drug-susceptible isolate from Barrow County Georgia, USA) were collected and made into a slurry with water, filtered through 425- and 180-µm sieves, and centrifuged at 2500 rpm for 5 min, after which the supernatant was discarded. Kaolin (Sigma-Aldrich, St. Louis, MO) was then added and resuspended in sodium nitrate (SPG 1.25–1.3) (Feca-Med®; Vedco, Inc. St Joseph, MO, USA). The tube was then centrifuged at 2500 rpm for 5 min, after which the supernatant was passed through a 30-µm sieve and rinsed with distilled water, and reduced to a volume of 10 to 15 mL. The volume was adjusted to one egg per 5 µl using distilled water. The eggs were stored at room temperature for 2 h before placing them onto the Whatman® FTA® cards. Eggs were also placed onto Nematode Growth Medium (NGM) plates (Sulston and Hodgkin, 1988) and incubated at 26°C to obtain the first-stage (L1) larvae. After 48 h, larvae were rinsed off the plate with distilled water and centrifuged at 1000 rpm for 5 min. Larvae were counted, and the concentration adjusted to one larva per 5 µl. The larvae were stored at room temperature for 2 h before placing them onto the Whatman® FTA® cards. To obtain third-stage (L3) larvae, eggs were isolated from fresh feces from a research purpose-bred laboratory beagle (University of Georgia, AUP A2017 10-016-Y1-A0) infected with the Worthy isolate (Worthy 3.1F3Pyr; multiple-drug resistant isolate originally isolated from a greyhound dog, Florida, USA). Eggs were placed onto NGM plates and incubated at 26°C. After 7 days, larvae were rinsed off the plate with distilled water and centrifuged at 1000 rpm for 5 min. Larvae were counted, and the concentration adjusted to one larva per 5 µl. The larvae were stored at room temperature for 2 h before placing them onto the Whatman® FTA® cards.

Ascaridia dissimilis: Eggs of *A. dissimilis* (Isolate Wi: North Carolina, USA) were isolated from excreta of experimentally infected turkeys. Water was added to the excreta and made into a slurry, which was filtered using a 425-µm and 180-µm sieve to remove large debris. The remaining particulates were placed into 50-mL centrifuge tubes and centrifuged at 433*g* for 7 min. Supernatant was removed, and the pellet was resuspended in a saturated sucrose solution with a specific gravity of 1.15. The suspension was centrifuged as before, and eggs were isolated from the top layer. Eggs were rinsed over a 20-µm sieve with water to remove residual sucrose and then concentrated to one egg per 5 µl using deionized water. Multiple 5-µl aliquots of the egg solution were dispensed using a micropipette onto the Whatman® FTA® card.

Dirofilaria immitis: Blood was taken from a dog infected with the macrocyclic lactone (ML)-resistant Yazoo strain (Yazoo: originally isolated from a dog in Yazoo City, Mississippi, USA; see Maclean et al. (2017) for complete history). To obtain microfilariae, blood was collected in heparin tubes and centrifuged for 30 min at 2500 rpm, after which the supernatant was discarded. The pellet was suspended in 3.8% sodium citrate (Sigma-Aldrich, St. Louis, MO) and 15% saponin (Sigma-Aldrich) was added in a 1:7 dilution. The tube was then vortexed and centrifuged for 30 min at 2500 rpm, after which the supernatant was discarded and the pellet resuspended in 3.8% sodium citrate to the original blood volume, vortexed, and then centrifuged for 4 min at 2500 rpm. The pellet was then resuspended and mixed in a 1:9 solution of 10X phosphate-buffered saline (PBS) (Thermo Fisher Scientific, Waltham, MA) and distilled water. The tube was then centrifuged for 4 min at 2500 rpm, and the pellet resuspended in PBS. The microfilariae were then counted and adjusted accordingly to have one microfilaria per 5 µl and stored at room temperature for 2 h before placing them onto the Whatman® FTA® cards.

Dracunculus medinensis: Individual L1 samples were obtained as progeny of an adult female worm manually extracted from an infected dog in Tarangara village, Chad (9.068611 N, 18.708611 E) in 2016. This extraction forms part of the standard containment and treatment procedure for Guinea worm infections, as agreed upon and sanctioned by the World Health Organization and country ministries of health. The adult worm was submerged in ethanol in a microcentrifuge tube for storage; L1 stage progeny that were found settled on the bottom of the tube were collected for analysis.

Haemonchus contortus: Eggs representing the F5 generation of a genetic cross (described in Doyle et al. (2018)) were collected from fresh feces from experimentally infected sheep housed at the Moredun Research Institute, UK. All experimental procedures were examined and approved by the Moredun Research Institute Experiments and Ethics Committee and were conducted under approved UK Home Office licenses (PPL 60/03899) in accordance with the Animals (Scientific Procedures) Act of 1986. Briefly, feces were mixed with tap water and passed through a 210-µm sieve, then centrifuged at 2500 rpm for 5 min in polyallomer tubes. The supernatant was discarded, before adding kaolin to the fecal pellet, vortexing, and resuspending in a saturated salt solution. After centrifugation at 1000 rpm for 10 min, the polyallomer tube was clamped to isolate eggs, which were collected on a 38-µm sieve and rinsed thoroughly with tap water. Eggs were incubated on NGM plates at 20°C for 48 h to hatch to L1 stage larvae. In addition to freshly collected material, eggs collected in the same manner then stored at −20°C, from a previous generation of the cross, were also tested. Eggs and L1 larvae were resuspended in PBS and spotted onto Whatman FTA cards in 3 µl per egg/L1.

Schistosoma mansoni: Three collections of *S. mansoni* samples were used in this work. The first were field samples collected from humans on Lake Victoria fishing villages in Uganda as part of the LaVIISWA trial (Sanya et al., 2018). Ethical approval for this trail was given by the Uganda Virus Research Institute (reference number GC127), the Uganda National Council for Science and Technology (reference number HS 1183), and the London School of Hygiene & Tropical Medicine (reference number 6187). Parasite eggs were collected from participants' stool samples using a Pitchford-Visser funnel, washed with mineral water until clean, and transferred into a petri dish with water to be hatched in direct sunlight. After hatching, the miracidia were picked in 2-μl water using a pipette and placed on a Whatman® FTA® card for storage. The second were field samples collected as part of a repeated cross-sectional study of MDA exposure in school children in Uganda. Patient enrolment, including written consent, and sample collection have been described previously (Crellen et al., 2016). Ethical approvals for this study were granted by the Uganda National Council of Science and Technology (MoU sections 1.4, 1.5, 1.6) and the Imperial College Research Ethics Committee (EC NO: 03.36. R&D No: 03/SB/033E). Host stool was sampled 1 to 3 days prior to treatment with praziquantel (40 mg/kg) and albendazole (400 mg). A Pitchford-Visser funnel was used to wash and filter stool to retain parasite eggs. The filtrate was kept overnight in water and hatched the following morning in sunlight. Individual miracidia were isolated with a 20-μL pipette and transferred into petri dishes of nuclease-free water twice before spotting onto Whatman® FTA® cards. The third source of miracidia was derived from the livers of experimentally infected mice kept to maintain the *S. mansoni* life cycle at Wellcome Sanger Institute. Mouse infection protocols were approved by the Animal Welfare and Ethical Review Body (AWERB) of the Wellcome Sanger Institute, and in accordance with the UK Home Office approved project license P77E8A062. The AWERB is constituted as required by the UK Animals (Scientific Procedures) Act 1986 Amendment Regulations 2012. Balb/C mice (6–8 weeks old) were infected with 250 cercariae, after which livers were collected on day 40 postinfection. Eggs were isolated from the liver tissues using collagenase digestion followed by percoll gradient, and were washed well with sterile PBS, before being hatched in sterile conditioned water. The hatched individual miracidia were spotted onto Whatman® FTA® cards.

Strongyloides stercoralis: The *S. stercoralis* UPD strain and the isofemale isolate PVO1 were maintained in purpose-bred, prednisone-treated mix breed dogs according to protocol 804883 approved by the University of Pennsylvania Institutional Animal Care and Use Committee (IACUC), USA. IACUC-approved research protocols and all routine husbandry care of the animals were conducted in strict accordance with the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health, USA. Feces were collected, moisturized, and mixed with equal volume of charcoal and cultured at 22°C in 10-cm plates (Lok, 2007). Post-parasitic stage one and two larvae (L1/L2), free-living male and female adults, and infective third-stage (L3) were isolated by the Baermann technique after 24 and 48 h, and 6 days in these charcoal coprocultures, respectively (Lok, 2007; Jaleta et al., 2017). For this study, free living females (FL), L1 and L3 larvae were collected from the Baermann funnel sediments and washed three times using PBS.

Trichuris muris: Infection and maintenance of *T. muris* was conducted as described (Wakelin, 1967). The care and use of mice were in accordance with the UK Home Office regulations (UK Animals Scientific Procedures Act 1986) under the Project license P77E8A062 and were approved by the institutional AWERB. Female SCID mice (6–10 wk old) were orally infected under anaesthesia with isoflurane with a high dose (n = 400) of embryonated eggs from *T. muris* E-isolate. Mice were monitored daily for general condition and weight loss. At day 35 postinfection, mice were killed by exsanguination under terminal anesthesia, after which adult worms were harvested from cecums. Adult worms were cultured in RPMI 1640 supplemented with 10% fetal calf serum (v/v), 2 mM L-glutamine, penicillin (100 U/mL), and streptomycin (100 mg/mL; all Invitrogen), for 4 h or overnight, and eggs were collected. The eggs were allowed to embryonate for at least 6 weeks in distilled water, and infectivity was established by worm burden in SCID mice. *T. muris* eggs were hatched to produce sterile L1 larvae using 32% sodium hypochlorite in sterile water for 2 h at 37°C with 5% CO_2. Eggs were washed with RPMI 1640 supplemented with 10% fetal calf serum (v/v), 2 mM L-glutamine, penicillin (100 U/mL), and streptomycin (100 mg/mL; all Invitrogen), and incubated at 37°C with 5% CO_2 for 4 to 5 days until they hatched.

For each species, unless otherwise stated, pools of individuals were washed in sterile PBS, before being transferred to a petri dish. Individuals were identified under the microscope, after which 5-μl PBS containing an individual parasite was transferred onto a Whatman® FTA® card and dried for a minimum of 20 min at room temperature prior to storage or shipping to the Wellcome Sanger Institute, UK. The Whatman® FTA® cards with samples spotted were stored in a clean plastic bag in the dark at room temperature prior to analysis.

DNA Extraction

Five DNA extraction methods were tested for their ability to isolate and purify DNA compatible with whole-genome sequencing approaches. The choice of approaches was not systematic or comprehensive, but was available to us at the Wellcome Sanger Institute as low-input DNA extraction approaches that had the potential for high-throughput and/or automatic extraction protocols to generate sequencing libraries suitable for whole-genome sequencing. Despite the arbitrary choice of extraction kits, the comparison of multiple kits with multiple helminth species provides a unique data set that is not easily achieved at this scale outside of a research and development environment, and forms the basis for further comparison with other low-input kits as they become available.

To extract DNA, the sample spots on FTA cards were punched out manually into 96-well plates using either a Harris Punch or autonomously using robotics. The DNA extraction was carried out for each method as described below:

1. *Nexttec* (NXT): Extraction using the Nexttec 1-step DNA Isolation Kit for Tissues & Cells (cat: 10N.904; Waendel Technology Limited, UK) was performed according to manufacturer's guidelines. Proteinase lysis buffer (75 μl) was used for digestion.

2. *Bloodspot* (BSP): The bloodspot extraction was derived from the QIAamp DNA Investigator Kit (cat: 51104/51106; Qiagen), following the "DNA Purification from Dried Blood Spots (QIAamp DNA Mini Kit)" protocol according to manufacturer's guidelines.

3. *CGP* (CGP): Extraction using the CGP protocol (Moore et al., 2018) involved adding 30 µl of lysis buffer (1.25 µg/mL of Protease reagent (Qiagen; cat 19155) in Tris HCl pH 8.0, 0.5% Tween 20, 0.5% NP40) to individual Whatman® FTA® punches in a 96-well PCR plate. The samples were incubated at 50°C for an hour followed by protease inactivation by heating the samples to 75°C for 30 min.

4. *ForensicGem* (FGM): Extraction using the ForensicGEM Universal DNA extraction kit (cat: FUN0100; ZyGem) was performed according to the manufacturer's guidelines.

5. *PicoPure* (PIP): Extraction using the ARCTURUS® PicoPure® DNA Extraction Kit (cat: KIT0103; ThermoFisher) was performed according to the manufacturer's guidelines, using 75 µl of the extraction buffer.

When the extracted DNA from all the above methods was present in a volume greater than 25 µl, samples were cleaned with Agencourt AMPure XP beads (Beckman-Coulter) and eluted in 25-µl nuclease-free H$_2$O. The entire DNA samples were used downstream to make sequencing libraries. A summary of the number of species, life stages, and conditions tested, is presented in **Table 1**. The differences in the number of samples per species and per assay were largely dependent on the samples available to us, sampled across different times, and sometimes, sampled for different purposes, for example, while most test conditions contained between 6 and 10 replicates, the large number of *D. medinensis* samples (n = 129 for one condition) were prepared specifically for a different study but we have included it for comparative purposes, whereas the *S. mansoni* samples (n = 168 for five conditions) were readily available in-house at the

Wellcome Sanger Institute, and thus were first used for further validation before samples from other species that required collection from live animal hosts.

Library Preparation and Sequencing

DNA sequencing libraries for all samples were prepared using a protocol designed for library preparation of laser capture microdissected biopsy (LCMB) samples using the Ultra II FS enzyme (New England Biolabs) for DNA fragmentation as previously described (Lee-Six et al., 2019). A total of 12 cycles of PCR were used (unless otherwise stated in **Table S1**) to amplify libraries and to add a unique 8-base index sequence for sample multiplexing. Prior to sequencing, library concentration was determined using a fluorometric dsDNA quantification assay (AccuClear® Ultra High Sensitivity dsDNA Quantitation Kit; Biotium) following the manufacturer's instructions, and measured using a FLUOstar Omega fluorescence plate reader (BMG Labtech). These data were used to normalized the concentration of library DNA for multiplexing.

Multiplexed libraries were sequenced using the Illumina MiSeq platform with V2 chemistry 150 bp paired end (PE) reads. The *D. medinensis* samples were sequenced as part of a different study using the HiSeq 2500 platform with V4 chemistry 125 bp PE reads. Metadata for each sample, including sample IDs, sequencing lane IDs, ENA sample accession numbers, and data generated are described in **Table S1**. Raw sequence data used in this study is available under the European Nucleotide Archive (ENA; https://www.ebi.ac.uk/ena) study ID ERP114942.

Analysis

We performed sufficient low coverage sequencing on each sample to enable us to identify: (i) the proportion of on-target mapped reads, (ii) the proportion of duplicate reads, i.e. library artefacts, and (iii) the proportion of off-target contaminant reads. On-target reads are defined as reads that mapped to the genome

TABLE 1 | Summary of the species, life stages, and conditions tested using one or more of the DNA extraction approaches followed by whole-genome sequencing.

| Species | Species ID | Life stage | Life stage ID | Samples sequenced per DNA extraction approach | | | | |
				BSP	CGP	FGM	NXT	PIP
Haemonchus contortus	HC	Egg	EGG	10	10	4	10	10
	HC	Egg (frozen)	EGGf	6	6	0	6	6
	HC	Larval stage L1	LS1	10	10	4	10	10
Schistosoma mansoni	SM	Miracidia	MIR	59	36	10	51	12
Ancylostoma caninum	AC	Egg	EGG	0	8	0	0	0
	AC	Larval stage L1	LS1	0	8	0	0	0
	AC	Larval stage L3	LS3	0	8	0	0	0
Ascaridia dissimilis	AD	Egg	EGG	0	8	0	0	0
Dirofilaria immitis	DI	Microfilaria	MFL	0	8	0	0	0
Strongyloides stercoralis	SS	Free living females	FL	0	8	0	0	0
	SS	Larval stage L1	LS1	0	8	0	0	0
	SS	Larval stage L3	LS3	0	8	0	0	0
Trichuris muris	TM	Egg	EGG	0	8	0	0	0
	TM	Egg (bleached)	EGGb	0	8	0	0	0
	TM	Larval stage L1	LS1	0	8	0	0	0
Dracunculus medinensis	DM	Larval stage L1	LS1	0	0	0	129	0
Total samples per kit (n = 497 total)				85	150	18	206	38

of the species from which DNA extraction was performed, and thus, is a proxy for the specificity of the experiment. Similarly off-target reads are defined as sequencing reads that do not map to the genome of the species from which DNA was extracted, but are derived from putative contaminants. The presence and proportion of contaminant-derived sequencing reads in each sequencing library was further analyzed using the kmer-based classification tool, kraken (v0.10.6-a2d113dc8f) (Wood and Salzberg, 2014), which we have used to assign taxonomic labels to raw sequencing data. Kmers from raw reads are compared with a kraken database (custom database built at WSI containing human [GRCh38] and mouse [GRCm38] genomes, all plasmid, bacterial, and viral genomes, as well as all Illumina adapters, from NCBI at the date of generation [v.pi_qc_2015521]); any raw sequencing reads that match kmers present in the database are defined as contaminants, whereas reads that do not match the database are defined as "kraken_unassigned." As no helminth genomes are present in the kraken database, these reads of interest will be in the kraken_unassigned category, for example, a high proportion of "kraken_unassigned" reads are indicative of low contamination (or at least low known contaminants present in the kraken database).

Reference genomes from each of the test species were obtained from the helminth genome repository WormBase Parasite (Howe et al., 2017) Release 12. Raw sequence data for each species were mapped to their respective reference genome (which included the mitochondrial genome) using *bwa* (v0.7.17-r1188) (Li, 2013) *mem* using default parameters (with the inclusion of -Y and -M options to use soft clipping for supplementary alignments, and to mark shorter hits as secondary, respectively), after which duplicate reads were marked using *Picard* (v2.5.0; https://github.com/broadinstitute/picard). *Samtools* (v1.3) flagstat and *bamtools* (v 2.3.0) stats were used to characterise the outcome of the mapping, the results of which were collated using *MultiQC* v1.3 (Ewels et al., 2016). Data were manipulated and visualized in the *R* (v3.5.0) environment

using the following packages: *ggplot2* (https://ggplot2.tidyverse.org/), *patchwork* (https://github.com/thomasp85/patchwork), and *dplyr* (https://dplyr.tidyverse.org/).

The relative ratio of the mitochondrial to nuclear genome was calculated using *bedtools* (v2.17.0) (Quinlan and Hall, 2010) *makewindows* and *samtools* (v1.6) *bedcov*. This was performed for each life stage of each species described in **Table 2**, whereby average read coverage from the mapped reads (bam files) were determined for the autosomes (characterised as defined chromosomes for *S. mansoni* and *H. contortus*, otherwise, the longest 10 scaffolds for the remaining species were used) and mitochondrial genomes. These data were used to compare the theoretical multiplexing that could be performed per species, either for whole-genome sequencing to achieve 30× genome-wide coverage, or alternatively, sufficient sequencing to achieve 100× coverage of the mitochondrial genome (see **Table 2** for results). Here, these data were calculated based on Illumina high-throughput sequencing using the NovaSeq 6000 system with S4 2 × 150 bp PE chemistry generating 2.5 terabases (Tb) of data and accounting for 85% mapping rate. However, to aid in the design of experiments for other species and sequencing platforms, the number of samples that can be multiplexed can be determined by:

$$
Multiplex\left(WGS\right)
$$
$$
= \frac{Sequencer\ lane\ capacity\left(bp\right)}{\left[\ genome\ size\left(bp\right)\ \right] \times \left[\ target\ genome\ coverage\ \right] \times \left[\ \dfrac{100}{average\ mapping\ efficiency\left(\%\right)}\ \right]}
$$

Where the "genome size" is the estimated number of base pairs in the genome, the "targeted genome coverage" is the intended number of reads covering every base pair of the genome, and the "average mapping efficiency" is the typical proportion of mapped reads from a sequencing library. This latter metric is difficult to predict before performing any

TABLE 2 | Breakdown of sequencing strategies per species based on whole-genome sequencing at 30× coverage and whole-genome sequencing to achieve 100× whole mitochondrial genome coverage.

Species / stage	Genome assembly size [1]	Number of samples that can be multiplexed [2]	mtDNA / nuclear ratio [3]	Sampled multiplexed targeting mtDNA at 100× coverage [2]	Nuclear genome coverage per 100× mtDNA genome
Schistosoma mansoni miracidia	409	173	142.77	7418	0.70
Haemonchus contortus egg	283	250	70.76	5313	1.41
Haemonchus contortus L1	283	250	44.55	3345	2.24
Dracunculus medinensis L1	103	688	70.81	14608	1.41
Dirofilaria immitis microfilaria	88	805	14.27	3446	7.01
Strongyloides stercoralis FL	42	1687	11.06	5597	9.04

[1]Genome size: Obtained from Wormbase Parasite v12.

[2]Multiplex: total number of samples per NovaSeq 6000 sequencing run with S4 2 × 150 bp PE chemistry generating 2.5 Tb of data and accounting for 85% mapping rate. Theoretically possible, but may be limited by barcoding available.

[3]mtDNA/nuclear ratio: based on coverage of mitochondrial and nuclear-derived sequencing reads, normalized to the nuclear genome coverage.

sequencing, however, knowledge of this from prior sequencing experiments can improve the actual coverage achieved by applying this scaling factor.

This equation can be extended to low-coverage whole-genome sequencing to target a defined mitochondrial DNA coverage using the following:

$$Multiplex\left(mtDNA\right)$$
$$= \frac{Sequencer\ lane\ capacity\left(bp\right)}{\left[\mathbf{genome\ size}\left(bp\right)\right]\times\left[\frac{1}{mtDNA\ to\ nuclear\ ratio}\right]\times\left[\frac{100}{average\ mapping\ efficiency\left(\%\right)}\right]\times\left[target\ mtDNA\ coverage\right]}$$

where "mtDNA to nuclear ratio" is the number of mitochondrial genomes per nuclear genome, which can be determined by relative reads counts performed here, or by alternative molecular approaches, such as qPCR.
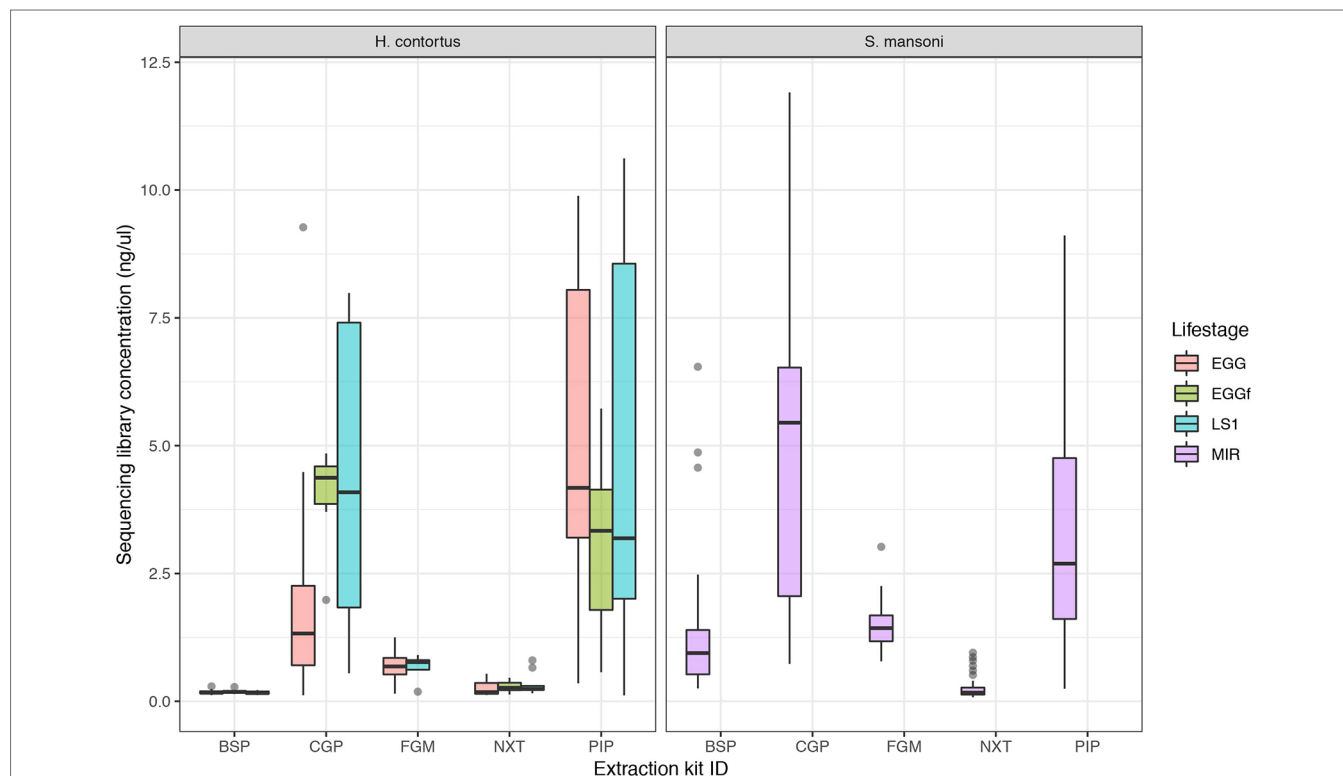
The code to reproduce the analysis and figures for this manuscript is described in https://github.com/stephenrdoyle/helminth_extraction_wgs_test.

RESULTS AND DISCUSSION

The aim of this work was to determine the feasibility and efficiency of using a low input DNA extraction and library preparation

approach for whole-genome sequencing of individual egg and larval stages of helminth parasites. We have targeted immature life stages that are found in the environment for which it is possible to collect samples non-invasively. We first tested our approaches on the nematode *H. contortus* and the trematode *S. mansoni*. Five approaches were tested using single egg (fresh and frozen) and L1 of *H. contortus*, and miracidia of *S. mansoni*. Attempts to quantify the raw DNA extractions were inconsistent and largely unsuccessful, due to the extremely low DNA yield per extraction; in the life stages tested, we expect this yield to be in the range of 10 to 100 s of pg of DNA per sample. Quantification of the sequencing libraries of these extractions did however yield sufficient DNA, and revealed significantly more DNA recovered using the CGP and PIP extractions than the other three protocols (**Figure 1**). Although not a direct measure, sequencing library concentration is likely a sufficient proxy for DNA extraction efficiency, given that the library preparation was standardized across all samples.

We determined the success of the library preparation protocols by comparing (i) the proportion of reads mapped to the genome, representing "on-target" mapping as a measure of specificity (**Table S1**: mapped_reads_percent); (ii) the proportion of reads that matched a "contamination" database, which did not contain helminth genome sequences, and thus represented DNA derived from "off-target" sources such as host
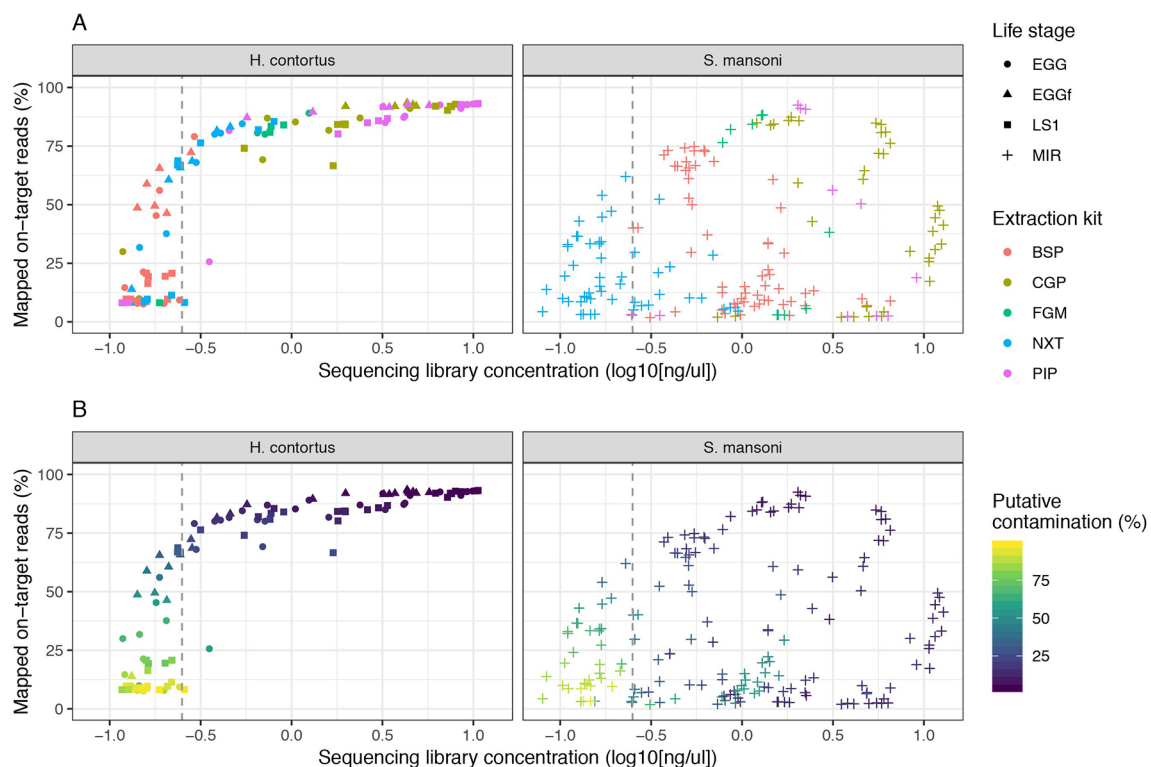
FIGURE 1 | Comparison of sequencing library concentration (ng/μl) as a proxy for relative DNA extraction efficiency for the five DNA extraction protocols tested. DNA from multiple life stages were extracted, including from freshly collected egg (EGG), frozen eggs (EGGf), and L1 stage larvae (LS1) of *Haemonchus contortus*, and miracidia (MIR) of *Schistosoma mansoni*. The number of samples in each comparison is presented in **Table 1**, and the raw data used are described in **Table S1**.

or bacterial species commonly present in these samples; those reads that do not match the contamination database are defined as "kraken_unassigned" and are putative helminth-derived reads (**Table S1**: kraken_unassigned_percent); and (iii) the proportion of duplicate reads, which typically represent library preparation artefacts due to over-amplification of DNA during PCR (**Table S1**: duplicate_reads_percent).

We examined the impact of library strength on mapping frequency for all DNA extraction protocols tested (**Figure 2**). For *H. contortus*, a clear correlation between these two variables were observed, with an inflection point at approximately 0.25 ng/μl, below which the proportion of reads mapped rapidly decreased toward 0 (**Figure 2A**). Similarly, below this point, the proportion of reads classified as contamination increased in frequency (**Figure 2B**). There was, however, a distinct difference between *H. contortus* and *S. mansoni* in the overall proportion of reads mapped, and the frequency of contaminating reads, with greater variation in both parameters in the *S. mansoni* samples. While some of this variation may reflect differences in extracting DNA from the two species, the majority of *S. mansoni* MIR samples were isolated and aliquoted onto FTA cards under less clean conditions in the field and stored for between 2 and 5 years before processing, as compared with the

laboratory prepared *H. contortus* samples that were collected and prepared directly from a current infection (EGG and LS1) or within the previous 6 months (EGGf) of processing. We achieved some on-target mapping for all extraction kits tested (**Figure S1A**); however, significant variation was observed between approaches. BSP and NXT generally performed poorly, with either low (median = 19.42; median absolute deviation [MAD] = 16.50) or significant variance (median = 66.20; MAD = 26.18) in mapping frequency between samples observed, respectively. PIP performed consistently well with high mapping rates across all stages in *H. contortus* (median = 90.30; MAD = 3.96); however, it was poor in *S. mansoni* (median = 11.00; MAD = 12.73). The duplication rate of all conditions were within an acceptable low range (median = 0.46, MAD = 0.4), with only 2% of samples having greater than 5% duplicate reads (**Figure S1B**). This suggests that the low DNA input did not noticeably impact duplication rates, and therefore, with greater sequencing depth of the same sequencing libraries, we would expect unbiased genome-wide coverage (at least relative to sequencing of libraries derived from higher DNA input). CGP and FGM performed most consistently between stages and species; however, FGM had higher variance and duplication rates relative to CGP across all samples tested. Considering the
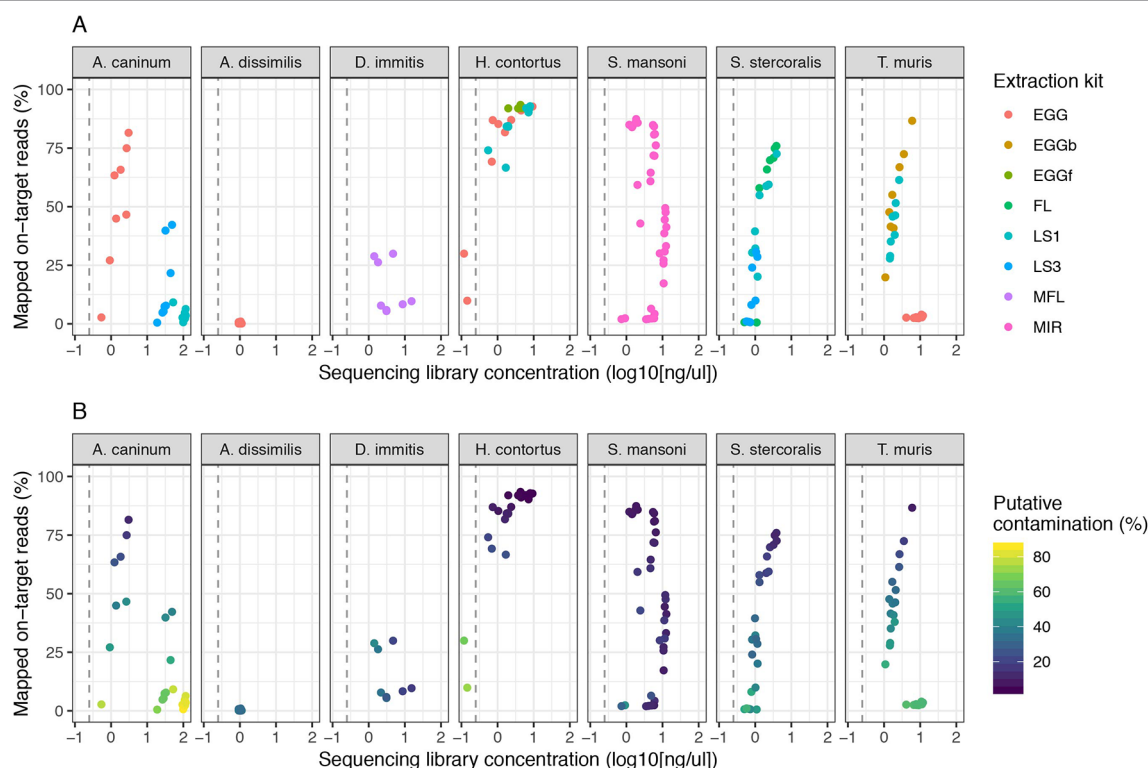
FIGURE 2 | Impact of sequencing library concentration on mapped on-target reads for five low-input DNA extraction protocols. Each point represents an individual sequencing library sample, colored by extraction protocol **(A)** or proportion of putative contaminating reads (100 - percent_unclassified_reads) **(B)**. The dashed grey vertical line represents a library concentration of 0.25 ng/μl. Extraction protocols were performed on *Haemonchus contortus* eggs (EGG), frozen eggs (EGGf), and L1 (LS1) stages, and *Schistosoma mansoni* miracidia (MIR). The number of samples in each comparison is presented in **Table 1**, and the raw data used are described in **Table S1**.

higher library strength (suggestive of greater DNA recovery efficiency), consistent mapping efficiency, and the cost effective protocol, we chose the CGP extraction to explore further.

We expanded our analysis of the CGP protocol to a total of seven helminth species for which samples were available, including a total of five distinct life stages (**Figure 3**). High variability in mapping was observed between species, with 50% or greater mapping frequency achieved in at least one life stage of five of the seven species tested (**Figure 3A**, **Figure S2A**). Clear differences were observed between multiple life stages tested within a species, likely reflecting differences in extraction efficiency per life stage, for example: for *A. canium*, reads from eggs (median = 54.98) mapped much more effectively than L1 (median = 3.61) or L3 stages (median = 7.57), and in *S. stercoralis,* full-length females (median = 67.84) and L1 (median = 47.19) performed better than L3 (median = 9.01). The proportion of contaminating reads tended to increase as library concentration decreased for all species (**Figure 3B**); however, some samples for which mapping was poor but higher libraries concentrations were obtained were highly enriched for contaminants, for example, *A. canium* L1 larvae and *T. muris* eggs. Interestingly, *T. muris* L1 larvae (median = 41.85) and bleached eggs (median = 51.35) performed much better than untreated eggs (median = 2.72); bleaching is experimentally used for promoting hatching of *T. muris*, by dissolving the egg shell layers and in turn improving access to DNA within. However, bleached eggs were embryonated and developmentally more advanced than untreated unembryonated eggs, and therefore would have more DNA available for library preparation. Similar to untreated *T. muris* eggs, *A. dissimilis* eggs performed poorly (median = 0.16); for both species, few if any nematode sequences were recovered and the majority of sequencing reads were contaminating bacterial-derived contaminants, perhaps indicative of the challenge of accessing material with the environmentally resistant egg. Analysis of *A. dissimilis* was further limited by the lack of a reference genome for this species; we mapped against the available genome of *Ascaris lumbricoides*, the nearest species for which a reference was available, and therefore at best, would have expected suboptimal mapping due to sequence divergence. The duplication rates remain low for all species and conditions tested (**Figure S2B**), consistent with our initial observation that library input did not significantly influence duplication rates (**Figure S1B**).

We also generated data from microfilaria of *D. medinensis*; these samples were extracted using NXT rather than CGP as part of a separate study; however, we have included these for comparison (**Figure S3**). High mapping rates were observed for almost all libraries generated; however, the duplication rates per

FIGURE 3 | Impact of sequencing library concentration on mapped on-target reads from multiple life stages of seven helminth species. Each point represents an individual sequencing library samples, colored by life stage **(A)** or proportion of putative contaminating reads (100 - percent_unclassified_reads) **(B)**. DNA was extracted from all samples using the CGP protocol. Consistent with **Figure 2**, the dashed grey vertical line represents 0.25 ng/μl. Egg, untreated egg; EGGb, bleached egg; EGGf, frozen egg; FL, full-length female; LS1, larval stage L1; LS3, larval stage L3; MFL, microfilaria; MIR, miracidia. The number of samples in each comparison is presented in **Table 1**.

library were also high (median = 36.90); this was unexpected, given NXT did not produce high duplication rates under the initial conditions tested with *H. contortus* or *S. mansoni*, nor were there excessive PCR cycles used in this instance. Duplication rates were not correlated with library concentration.

In summary, we present successful DNA extraction followed by whole-genome sequencing of individual egg and larval stages from six of eight parasites species examined. These results significantly extend the possibility of genomic analyses for life stages for which, at best, were limited to low-resolution, low-throughput PCR based assays without the addition of whole-genome amplification. Whatman® FTA® cards provide a convenient substrate for sample collection and storage, and do not limit the application of direct DNA extraction and whole-genome sequencing of parasite samples, even for field samples as demonstrated for *S. mansoni* miracidia that were collected and processed in Uganda before they were transported to the UK. Further optimization is required to improve the DNA recovery from eggs, for example, from *A. dissimilis* and *T. muris*, to provide greater applicability of our approaches to species that generate particularly environmentally resistant stages, such as the soil transmitted helminths. The application of whole-genome sequencing to diagnose and monitor helminth infections *at scale* is largely limited by the costs of library preparation and sequencing, and therefore, will be restricted to niche applications of the technology. However, the use of low-coverage whole-genome sequencing with the specific aim to target the mitochondrial genome, which is present at a higher copy number than the nuclear genome (**Table 2**), may be a cost-effective alternative and potentially provides greater diagnostic information than low-throughput PCR-based diagnostics. Continued development of genomic technologies and the associated reduction in sequencing and library preparation costs will make screening large samples by genome sequencing more routine as in viral (Dudas et al., 2017) and bacterial (Domman et al., 2018) population studies. In doing so, the ability to derive high resolution data may provide insight into, for example, (i) within host reproductive dynamics, including within host population size, differential fecundity (Hildebrandt et al., 2014), and reproductive traits such as polyandry (Doyle et al., 2018), (ii) defining parasite transmission zones and rates of transmission between zones to prioritize treatment foci (Crawford et al., 2019; this special edition, in review), (iii) defining effective population sizes of parasite populations (Sallé et al., 2019), and using this to estimate the impact of control strategies over time, and/or (iv) discrimination between reintroduction and recrudescence of parasites in regions where parasite control has been successful (Koala et al., 2019). The use of genomics to provide information-rich data will be increasingly important for diagnostic and surveillance purposes broadly (Cotton et al., 2018), and will be particularly informative as efforts to control human infective helminths using MDA move from control to elimination.

DATA AVAILABILITY

The datasets generated for this study can be found in ENA, ERP114942.

ETHICS STATEMENT

Given a number of host/parasite systems are described, we have provided information about the individual ethical approvals obtained associated with each species in the "Sample Collection" subheading of the *Methods* section.

AUTHOR CONTRIBUTIONS

Conceptualization: SRD, JAC, NH; Methodology: SRD, GS, NH; Software: SRD; Formal analysis: SRD; Investigation: SRD, GS; Resources: FA, TJ, JL, JBC, PC, JW, TC, ET, MD-C, PE, RL, KM, CM, TM, BS, PO, JT, AW, ED, RK, DB, MB; Data curation: SRD, NH; Writing – original draft preparation: SRD; Writing – review and editing: All authors; Visualization: SRD; Supervision: SRD, JAC, NH; Project administration: NH.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00826/full#supplementary-material

TABLE S1 | Complete metadata per sample (excel workbook)

FIGURE S1 | Summary boxplots of mapped on target reads (A) and duplicate reads (B) for all extraction kits tested. The number of samples tested per kit, per life stage is shown at the top of (A).

FIGURE S2 | Summary boxplots for mapped on target reads (A) and duplicate reads (B) for all species tested using the CGP protocol. The number of samples tested per species, per life stage is shown at the top of (A).

FIGURE S3 | Analysis of *D. medinensis* using the NXT protocol. (A) Mapped "on-target" reads. (B) Duplicate reads. (C) Comparison of the effect of sequencing library concentration on mapping efficiency, colored by the proportion of duplicate reads. (D) Comparison of the effect of sequencing library concentration on mapping efficiency, colored by the proportion of putative contaminant reads identified using Kraken (100 - percent_unclassified_reads).

REFERENCES

Avramenko, R. W., Redman, E. M., Lewis, R., Yazwinski, T. A., Wasmuth, J. D., and Gilleard, J. S. (2015). Exploring the gastrointestinal 'nemabiome': deep amplicon sequencing to quantify the species composition of parasitic nematode communities. *PLoS One* 10(12), e0143559. doi: 10.1371/journal.pone.0143559

Avramenko, R. W., Redman, E. M., Melville, L., Bartley, Y., Wit, J., Queiroz, C., et al. (2019). Deep amplicon sequencing as a powerful new tool to screen for sequence polymorphisms associated with anthelmintic resistance in parasitic nematode populations. *Int. J. Parasitol.* 49 (1), 13–26. doi: 10.1016/j.ijpara.2018.10.005

Boué, F., El Berbri, I., Hormaz, V., Boucher, J.-M., El Mamy, A. B., Traore, A., et al. (2017). Use of FTA® Card methodology for sampling and molecular characterization of *Echinococcus granulosus* sensu lato in Africa. *Exp. Parasitol* 173, 29–33. doi: 10.1016/j.exppara.2016.12.016

Campbell, S. J., Stothard, J. R., O'Halloran, F., Sankey, D., Durant, T., Ombede, D. E., et al. (2017). Urogenital Schistosomiasis and soil-transmitted helminthiasis (STH) in Cameroon: an epidemiological update at Barombi Mbo and Barombi Kotto Crater Lakes assessing prospects for intensified control interventions. *Infect. Dis. Poverty* 6 (1), 49. doi: 10.1186/s40249-017-0264-8

Cotton, J. A., Berriman, M., Dalén, L., and Barnes, I. (2018). Eradication genomics—lessons for parasite control. *Science* 361 (6398), 130–131. doi: 10.1126/science.aar6609

Crawford, K. E., Hedtke, S. M., Doyle, S. R., Kuesel, A. C., Armoo, S., Osei-Atweneboana, M., et al. (2019). Utility of the Onchocerca volvulus mitochondrial genome for delineation of parasite transmission zones. *bioRxiv* 732446. doi: 10.1101/732446

Crellen, T., Walker, M., Lamberton, P. H. L., Kabatereine, N. B., Tukahebwa, E. M., Cotton, J. A., et al. (2016). Reduced efficacy of praziquantel against *Schistosoma mansoni* is associated with multiple rounds of mass drug administration. *Clin. Infect. Dis.* 63 (9), 1151–1159. doi: 10.1093/cid/ciw506

Domman, D., Chowdhury, F., Khan, A. I., Dorman, M. J., Mutreja, A., Uddin, M. I., et al. (2018). Defining endemic cholera at three levels of spatiotemporal resolution within Bangladesh. *Nat. Genet.* 50 (7), 951–955. doi: 10.1038/s41588-018-0150-8

Doyle, S. R., Laing, R., Bartley, D. J., Britton, C., Chaudhry, U., Gilleard, J. S., et al. (2018). A genome resequencing-based genetic map reveals the recombination landscape of an outbred parasitic nematode in the presence of polyploidy and polyandry. *Genome Biol. Evol.* 10 (2), 396–409. doi: 10.1093/gbe/evx269

Dudas, G., Carvalho, L. M., Bedford, T., Tatem, A. J., Baele, G., Faria, N. R., et al. (2017). Virus genomes reveal factors that spread and sustained the Ebola epidemic. *Nature* 544 (7650), 309–315. doi: 10.1038/nature22040

Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32 (19), 3047–3048. doi: 10.1093/bioinformatics/btw354

Gower, C. M., Shrivastava, J., Lamberton, P. H. L., Rollinson, D., Webster, B. L., Emery, A., et al. (2007). Development and application of an ethically and epidemiologically advantageous assay for the multi-locus microsatellite analysis of *Schistosoma mansoni*. *Parasitology* 134 (Pt 4), 523–536. doi: 10.1017/S0031182006001685

Hildebrandt, J. C., Eisenbarth, A., Renz, A., and Streit, A. (2014). Reproductive biology of *Onchocerca ochengi*, a nodule forming filarial nematode in Zebu cattle. *Vet. Parasitol.* 205 (1-2), 318–329. doi: 10.1016/j.vetpar.2014.06.006

Howe, K. L., Bolt, B. J., Shafie, M., Kersey, P., and Berriman, M. (2017). WormBase ParaSite—a comprehensive resource for helminth genomics. *Mol. Biochem. Parasitol.* 215, 2–10. doi: 10.1016/j.molbiopara.2016.11.005

Howick, V. M., Russell, A., Andrews, T., Heaton, H., Reid, A. J., Natarajan, K. N., et al. (2019). The Malaria Cell Atlas: a comprehensive reference of single parasite transcriptomes across the complete plasmodium life cycle. *Science* 365 (6455), eaaw2619. doi: 10.1126/science.aaw2619

Jaleta, T. G., Zhou, S., Bemm, F. M., Schär, F., Khieu, V., Muth, S., et al. (2017). Different but overlapping populations of *Strongyloides stercoralis* in dogs and humans-dogs as a possible source for zoonotic strongyloidiasis. *PLoS Negl. Trop. Dis.* 11 (8), e0005752. doi: 10.1371/journal.pntd.0005752

Koala, L., Nikièma, A. S., Paré, A. B., Drabo, F., Toé, L. D., Belem, A. M. G., et al. (2019). Entomological assessment of the transmission following recrudescence of onchocerciasis in the Comoé Valley, Burkina Faso. *Parasit. Vectors* 12 (1), 34. doi: 10.1186/s13071-019-3290-5

Le Clec'h, W., Chevalier, F. D., McDew-White, M., Allan, F., Webster, B. L., Gouvras, A. N., et al. (2018). Whole genome amplification and exome sequencing of archived *Schistosome miracidia*. *Parasitology* 145 (13), 1739–1747. doi: 10.1017/S0031182018000811

Lee-Six, H., Ellis, P., Osborne, R. J., Sanders, M. A., Moore, L., Georgakopoulos, N., et al. (2019). The landscape of somatic mutation in normal colorectal epithelial cells. *bioRxiv* 416800. doi: 10.1101/416800

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997 [q-bio.GN]*.

Lok, J. B. (2007). "Strongyloides stercoralis: a model for translational research on parasitic nematode biology." *The Online Review of C. Elegans Biology*. (WormBook) 1–18. doi: 10.1895/wormbook.1.134.1

Maclean, M. J., Savadelis, M. D., Coates, R., Dzimianski, M. T., Jones, C., Benbow, C., et al. (2017). Does evaluation of *in vitro* microfilarial motility reflect the resistance status of *dirofilaria immitis* isolates to macrocyclic lactones? *Parasit. Vectors* 10 (Suppl 2), 480. doi: 10.1186/s13071-017-2436-6

Marek, M., Zouhar, M., Douda, O., Maňasová, M., and Ryšánek, P. (2014). Exploitation of FTA cartridges for the sampling, long-term storage, and DNA-based analyses of plant-parasitic nematodes. *Phytopathology* 104 (3), 306–312. doi: 10.1094/PHYTO-03-13-0067-R

Moore, L., Leongamornlert, D., Coorens, T. H. H., Sanders, M. A., Ellis, P., Dawson, K., et al. (2018). The mutational landscape of normal human endometrial epithelium. *bioRxiv* 505685. doi: 10.1101/505685

Ngara, M., Palmkvist, M., Sagasser, S., Hjelmqvist, D., Björklund, Å K., Wahlgren, M., et al. (2018). Exploring parasite heterogeneity using single-cell RNA-Seq reveals a gene signature among sexual stage *Plasmodium falciparum* parasites. *Exp. Cell Res.* 371 (1), 130–138. doi: 10.1016/j.yexcr.2018.08.003

Platt, R. N., McDew-White, M., Le Clec'h, W., Chevalier, F. D., Allan, F., Emery, A. M., et al. (2019). Ancient hybridization and adaptive introgression of an invadolysin gene in schistosome parasites. *Mol. Biol. Evol.* 2019, *msz154*. doi: 10.1093/molbev/msz154

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26 (6), 841–842. doi: 10.1093/bioinformatics/btq033

Reid, A. J., Talman, A. M., Bennett, H. M., Gomes, A. R., Sanders, M. J., Illingworth, C. J. R., et al. (2018). Single-Cell RNA-Seq Reveals Hidden Transcriptional Variation in Malaria Parasites. *eLife* 7, e33105. doi: 10.7554/eLife.33105

Sabina, J., and Leamon, J. H. (2015). Bias in whole genome amplification: causes and considerations. *Methods Mol. Biol.* 1347, 15–41. doi: 10.1007/978-1-4939-2990-0_2

Sallé, G., Doyle, S. R., Cortet, J., Cabaret, J., Berriman, M., Holroyd, N., et al. (2019). The global diversity of the major parasitic nematode *Haemonchus contortus* is shaped by human intervention and climate. *bioRxiv* 450692. doi: 10.1101/450692

Sanya, R. E., Nkurununugi, G., Spaans, R. H., Nampijja, M., O'Hara, G., Kizindo, R., et al. (2018). The impact of intensive versus standard anthelminthic treatment on allergy-related outcomes, helminth infection intensity and helminth-related morbidity in lake victoria fishing communities, Uganda: results from the LaVIISWA Cluster Randomised Trial. *Clin. Infect. Dis.* 68 (10), 1665–1674. doi: 10.1093/cid/ciy761

Shortt, J. A., Card, D. C., Schield, D. R., Liu, Y., Zhong, B., Castoe, T. A., et al. (2017). Whole genome amplification and reduced-representation genome sequencing of *Schistosoma japonicum miracidia*. *PLoS Negl. Trop. Dis.* 11 (1), e0005292. doi: 10.1371/journal.pntd.0005292

Small, S. T., Labbé, F., Coulibaly, Y. I., Nutman, T. B., King, C. L., Serre, D., et al. (2018). Human migration and the spread of the nematode parasite *Wuchereria bancrofti*. *Mol. Biol. Evol.* doi: 10.1093/mo lbev/msz116

Sulston, J., and Hodgkin, J. (1988). "Methods," in *The Nematode Caenorhabditis Elegans*. Ed. W. B. Wood (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Trevino, S. G., Nkhoma, S. C., Nair, S., Daniel, B. J., Moncada, K., Khoswe, S., et al. (2017). High-resolution single-cell sequencing of malaria parasites. *Genome Biol. Evol.* 9 (12), 3373–3383. doi: 10.1093/gbe/evx256

Tsai, I. J., Hunt, M., Holroyd, N., Huckvale, T., Berriman, M., and Kikuchi, T. (2014). Summarizing specific profiles in illumina sequencing from whole-genome amplified DNA. *DNA Res.* 21 (3), 243–254. doi: 10.1093/dnares/dst054

Wakelin, D. (1967). Acquired immunity to *Trichuris muris* in the albino laboratory mouse. *Parasitology* 57 (3), 515–524. doi: 10.1017/S0031182000072395

Webster, B. L. (2009). Isolation and preservation of schistosome eggs and larvae in RNAlater® facilitates genetic profiling of individuals. *Parasit. Vectors* 2 (1), 50. doi: 10.1186/1756-3305-2-50

Webster, B. L., Emery, A. M., Webster, J. P., Gouvras, A., Garba, A., Diaw, O., et al. (2012). Genetic diversity within *Schistosoma haematobium*: DNA barcoding reveals two distinct groups. *PLoS Negl. Trop. Dis.* 6 (10), e1882. doi: 10.1371/journal.pntd.0001882

Wood, D. E., and Salzberg, S. L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* 15 (3), R46. doi: 10.1186/gb-2014-15-3-r46

Xiao, N., Remais, J. V., Brindley, P. J., Qiu, D.-C., Carlton, E. J., Li, R.-Z., et al. (2013). Approaches to genotyping individual miracidia of *Schistosoma japonicum*. *Parasitol. Res.* 112 (12), 3991–3999. doi: 10.1007/s00436-013-3587-9

frontiers
in Genetics

Check for
updates

A Case for Using Genomics and a Bioinformatics Pipeline to Develop Sensitive and Species-Specific PCR-Based Diagnostics for Soil-Transmitted Helminths

Jessica R. Grant[1]*, Nils Pilotte[1,2] and Steven A. Williams[1,2]

[1] Department of Biological Sciences, Smith College, Northampton, MA, United States, [2] Molecular and Cellular Biology Program, University of Massachusetts, Amherst, MA, United States

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. Although 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.

Keywords: soil-transmitted helminth, molecular diagnostics, DNA diagnostics, polymerase chain reaction (PCR), quantitative PCR

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 to 2.1 million disability-adjusted life years (DALYs) and US $7.5 billion to US $138.9 billion in loss of productivity (Bartsch et al., 2016; Kyu et al., 2018). Efforts are underway to eliminate STH, with the ambitious goal of controlling morbidity by the year 2020 (Becker et al., 2018; Uniting to Combat NTDs). Mass drug administrations (MDA) and water, sanitation and hygiene (WASH) programs across endemic countries are making headway (Hicks et al., 2015; Truscott et al., 2016; Weatherhead et al., 2017; Truscott et al., 2019), 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 (2018 Action Group

Meeting). 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 (Weatherhead et al., 2017).

Diagnostic techniques need to be inexpensive, practical, and give consistent results across technicians and laboratories. Importantly, they must be accurate, sensitive, and easily interpreted. Microscopy has long been relied on as the standard for diagnosis of intestinal parasites, including soil transmitted helminths (Beaver and Martin, 1968). Several copromicroscopic methods are in use, including FLOTAC (Cringoli et al., 2010), MINI-FLOTAC (Maurelli et al., 2014), several modifications of the McMaster technique (Mines, 1977), and Kato-Katz (Katz et al., 1972) (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 (Moser et al., 2018). With any of these tests, even highly trained microscopists can misidentify species or give inconsistent results (Krauth et al., 2012), and they are notoriously insensitive in regions with low infection rates (Nikolay et al., 2014; Buonfrate et al., 2015; Speich et al., 2015; Acosta Soto et al., 2017).

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 (Easton et al., 2016; Halfon et al., 2017; Holt et al., 2017). 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 (Andersen et al., 2013; Sarhan et al., 2015; Hidalgo et al., 2018; Papaiakovou et al., 2018). 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 (Harmon et al., 2006; Nunes et al., 2006; Kaisar et al., 2017). Although molecular diagnostics are not inexpensive, microscopy techniques are also expensive and can be difficult to scale up, whereas 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 (Easton et al., 2017; Papaiakovou et al., 2018; Truscott et al., 2019). Until recently, most PCR-based assays have targeted well-characterized and conserved regions, such as ITS and 18S (Verweij and Stensvold, 2014; Hii et al., 2018), but increased availability of whole-genome sequence data is facilitating the discovery of more sensitive and species-specific genomic targets (Pilotte et al., 2016a; Papaiakovou et al., 2017).

Repetitive elements are essential parts of eukaryotic genomes that have structural and regulatory functions (Shapiro and Sternberg, 2005; López-Flores and Garrido-Ramos, 2012), and different types of repetitive DNA elements have been studied and classified (Charlesworth et al., 1994; Plohl et al., 2008; López-Flores and Garrido-Ramos, 2012; Biscotti et al., 2015). Ribosomal DNA is found in repeat arrays (Lafontaine and Tollervey, 2001; Pruesse et al., 2007). These have traditionally been used for primer design and can give sensitive results depending on the size of the array (Verweij and Stensvold, 2014). However,

the repeat is oftentimes conserved between species and even genera, and rDNA-based assays are often less specific than those designed from other repeat types (Pilotte et al., 2016; O'Connell et al., 2018).

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 (López-Flores and Garrido-Ramos, 2012). Copy number can be quite variable in mini- and micro-satellites but larger satDNA monomers are more consistent within species (López-Flores and Garrido-Ramos, 2012). 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 (López-Flores and Garrido-Ramos, 2012).

Other types of repetitive DNA include transposons and retrotransposons, which are dispersed throughout the genome. These include short interspersed nuclear elements (SINEs) which are 100 to 500 bases long, and long interspersed nuclear elements (LINEs) which are larger—6,000 or 7,000 bases long. These may also be useful for repeat-based assay design (Funakoshi et al., 2017).

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 1**). Repetitive elements make up even greater percentages in other eukaryotes (Brindley et al., 2003; Wickstead et al., 2003; Shapiro and Sternberg, 2005; Coghlan et al., 2019), up to an astonishing 97% in some plants (Flavell et al., 1974; Piednoël et al., 2012). 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 species-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, although 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 (Ganley and Kobayashi, 2007; Teruel et al., 2014). 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 (Onyabe and Conn, 1999; Lindner and Banik, 2011) 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

TABLE 1 | Genome size of representative helminth species and the amount/percent of genome masked as repetitive. Low complexity repeats and simple (for example di- or tri-nucleotide repeats) are not targets for molecular assays.

Species	Assembly size (Mb)	Repeat-masked (Mb)	Repeat %	% of assembly that is low complexity/ simple repeats
Ancylostoma ceylanicum (Hii et al., 2018)	349	128.6	36.8	0.5
Ancylostoma duodenale (Hii et al., 2018)	332.9	116.1	34.9	0.7
Ascaris lumbricoides germline (Okorie and de Souza, 2016)	334	n/a	16.8	n/a
Ascaris lumbricoides soma (Okorie and de Souza, 2016)	291	n/a	7	1.5
Necator americanus (Masny et al., 2016; Hii et al., 2018)	244.1	67.1	27.5	1.2
Strongyloides stercoralis (Pilotte et al., 2017; Hii et al., 2018)	42.7	4.4	10.3	4.2
Trichuris trichiura (Hii et al., 2018)	75.5	18.4	24.4	0.4

found population-level copy number variation in ribosomal repeats of different species (Bik et al., 2013; Schaap et al., 2013; Mascagni et al., 2018; Zhao and Gibbons, 2018), 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 (Umemori et al., 2013). An understanding of copy number variation and its impact on assay sensitivity will likely need to be studied on an individual species-by-species 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 (McReynolds et al., 1986; Meredith et al., 1989; Chanteau et al., 1994; Nekrutenko et al., 2000; Hamburger et al., 2001; Rao et al., 2002; Demas et al., 2011; Lodh et al., 2016). 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.

FINDINGS
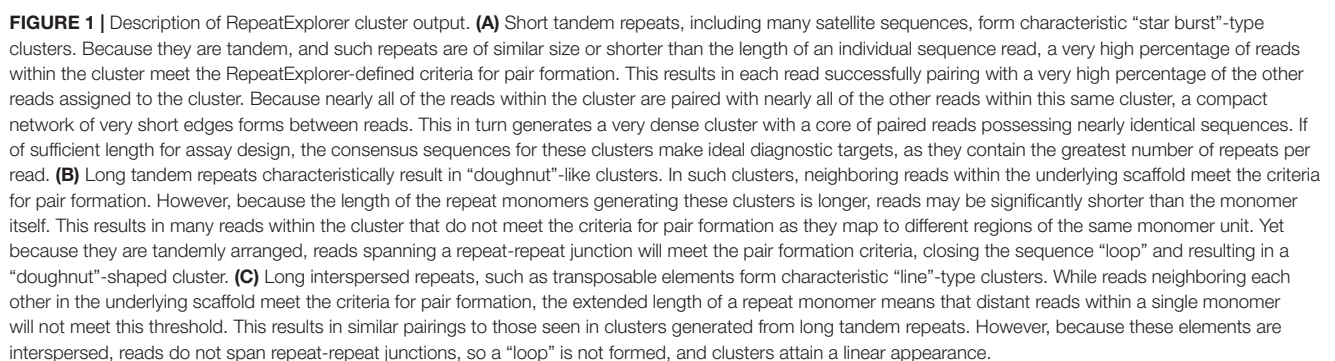
Diagnostics give an estimate of the true prevalence of a disease, with the probability of correctly estimating the truth given by the sensitivity and species-specificity of the diagnostic. WHO guidelines for when to treat a community are informed by the prevalence of disease in that community, and are 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 (Medley et al., 2016). Medley et al. (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 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 (Truscott et al., 2017; Ásbjörnsdóttir et al., 2017), which can only be accurately estimated with highly sensitive and species-specific diagnostics. Improved diagnostics are crucial to meet this need (Kongs et al., 2001; Andersen et al., 2013; Clarke et al., 2018; Hidalgo et al., 2018). 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 (Farrell and Anderson, 2018). 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 (Sanpool et al., 2012; Pilotte et al., 2016b; Ramírez et al., 2018; Zaky et al., 2018). Pooling of samples is a common way to decrease cost but diagnostic tests used for the screening of pooled samples need to be highly sensitive (Masny et al., 2016; Okorie and de Souza, 2016; Pilotte et al., 2017), 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 (Howe et al., 2016). Although 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 (Novák et al., 2010; Novák et al., 2013; Pilotte et al., 2016a). 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 1**), each read is represented by a node, and each sequence pair (by default defined as ≥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

FIGURE 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.

low-depth sequence data sets, low copy number sequences will not be well represented and will, therefore, graph as individual nodes or small clusters, whereas 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 copy number of the repeat 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 (Novák et al., 2017) 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 intragenomic variation. We have developed, and made available here, custom Python scripts to parse the output from RepeatExplorer and return highly repetitive sequences (https://github.com/JessicaGrant/RepeatTargetScripts). 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 (Pilotte et al., 2016a). Care is needed in choosing which repeat to select for use as a diagnostic target, as some may be found in closely related species (Williams et al., 2000; Rao et al., 2006); however, similar to what has been reported in the literature (Subirana and Messeguer, 2013; Subirana et al., 2015), 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 (Rao et al., 2006). 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 (Pickering et al., 2019), Bangladesh (Benjamin-Chung et al., 2019), Ethiopia, Uganda, Timor Leste (Papaiakovou et al., 2017), Thailand (O'Connell et al., 2018), Liberia (Fischer et al., 2018), 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 (Ásbjörnsdóttir et al., 2018). However, testing biological samples, whether for diagnosing individuals or getting an overview of the epidemiological environment of a region, involves a myriad of factors, such as unexpected or emerging parasites, zoonotic infections, and misleading material in the samples. Although 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 (Speich et al., 2015). 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 (Cushion et al., 1990). 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 (Benjamin-Chung et al., 2019) 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 (Pontes et al., 2003; Stensvold et al., 2006; Knopp et al., 2014; Pilotte et al., 2016a; Easton et al., 2017; Ng-Nguyen et al., 2017). 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 that targeted 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. Not one organism was found in the amplicon sequencing that could explain all of the false-positive results. What material in the samples 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 on 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 (Irisarri-Gutiérrez et al., 2016). Additional examples of misidentification of hookworm ova as other eggs (Ralph et al., 2006; Werneck et al., 2007; Yong et al., 2007) 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 (Llewellyn et al., 2016; Pilotte et al., 2016a). 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 (George et al., 2015; Chidambaram et al., 2017) 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. Our repeat-based *A. duodenale* assay, on the other hand, only detects *A. duodenale*, and so all of the *A. ceylanicum*-containing samples were negative. We have since used our pipeline to develop a species-specific qPCR assay for *A. ceylanicum* (Papaiakovou et al., 2017), which is more sensitive and specific than the ITS-based assay and easier to use than semi-nested PCR.

In a similar case, the same ITS-based primer set (Llewellyn et al., 2016) was used to investigate *Ancylostoma duodenale* in a field study of a refugee population in Thailand (O'Connell et al., 2018). 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 (Papaiakovou et al., 2017), 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, although 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. It also highlights that *A. ceylanicum* may be a much more common human pathogen than previously supposed. Here, the repeat-based, species-specific assays can be used to identify the true prevalence of various related parasite species.

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 *Trichuris suis*, which commonly infects pigs, and *Trichuris vulpis*, usually found in dogs, have also been found infecting humans (Areekul et al., 2010; Mohd-Shaharuddin et al., 2019). 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 (Pilotte et al., 2016a). 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. On the other hand, the use of the repeat-based, highly specific assays gives a true picture of the prevalence of various species.

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 (Fischer et al., 2018), 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 was not expected to be found 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. However, 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.

CONCLUSION

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 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 we think that the pipeline for assay development has potential for improving diagnostic sensitivity for many other classes of infectious agents.

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 antihelminthic resistance, emerging or zoonotic diseases, and optimal threshold levels for elimination of disease.

BIOINFORMATIC RESOURCES

RepeatExplorer (Novák et al., 2010; Novák et al., 2013) is made freely available here:

http://repeatexplorer.org/

We make our scripts available for others to adapt and use here:

https://github.com/JessicaGrant/RepeatTargetScripts

AUTHOR CONTRIBUTIONS

Conceptualization: JG, NP, and SW. Funding acquisition: SW. Writing—original draft: JG. Writing—review and editing: JG, NP, SW.

REFERENCES

Acosta Soto, L., Santísima-Trinidad, A. B., Bornay-Llinares, F. J., Martín González, M., Pascual Valero, J. A., and Ros Muñoz, M. (2017). Quantitative PCR and Digital PCR for Detection of Ascaris lumbricoides Eggs in Reclaimed Water. BioMed Res. Int. 2017, 1–9. doi: 10.1155/2017/7515409

Andersen, L. O., Röser, D., Nejsum, P., Nielsen, H. V., and Stensvold, C. R. (2013). Is Supplementary bead beating for DNA extraction from nematode eggs by use of the NucliSENS easyMag protocol necessary? J. Clin. Microbiol. 51, 1345–1347. doi: 10.1128/JCM.03353-12

Areekul, P., Putaporntip, C., Pattanawong, U., Sitthicharoenchai, P., and Jongwutiwes, S. (2010). Trichuris vulpis and T. trichiura infections among school children of a rural community in northwestern Thailand: the possible role of dogs in disease transmission. Asian Biomed. 4, 49–60. doi: 10.2478/abm-2010-0006

Ásbjörnsdóttir, K. H., Means, A. R., Werkman, M., and Walson, J. L. (2017). Prospects for elimination of soil-transmitted helminths. Curr. Opin. Infect. Dis. 30, 482. doi: 10.1097/QCO.0000000000000395

Ásbjörnsdóttir, K., Ajjampur, SSR, Anderson, R., Bailey, R., Gardiner, I., Halliday, K., et al. (2018). 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. 12(1):e0006166. doi: 10.1371/journal.pntd.0006166. eCollection

Bartsch, S. M., Hotez, P. J., Asti, L., Zapf, K. M., Bottazzi, M. E., Diemert, D. J., et al. (2016). The Global Economic and Health Burden of Human Hookworm Infection. PLoS Negl. Trop. Dis. 10, e0004922. doi: 10.1371/journal.pntd.0004922

Beaver, P. C., and Martin, L. K. (1968). Evaluation of Kato thick-smear technique for quantitative diagnosis of helminth infections *. Am. J. Trop. Med. Hyg. 17, 382–391. doi: 10.4269/ajtmh.1968.17.382

Becker, S. L., et al. (2018). Toward the 2020 goal of soil-transmitted helminthiasis control and elimination. PLoS Negl. Trop. Dis. 12, e0006606. doi: 10.1371/journal.pntd.0006606

Benjamin-Chung, J., Pilotte, N., Ercumen, A., Grant, J. R., Maasch, J. R. M. A., Gonzalez, A. M., et al. (2019). Comparison of multi-parallel qPCR and Kato-Katz for detection of soil-transmitted helminth infection among children in rural Bangladesh. bioRxiv 629501. doi: 10.1101/629501

Bik, H. M., Fournier, D., Sung, W., Bergeron, R. D., and Thomas, W. K. (2013). Intra-genomic variation in the ribosomal repeats of nematodes. PLOS ONE 8, e78230. doi: 10.1371/journal.pone.0078230

Biscotti, M. A., Olmo, E., and Heslop-Harrison, J. S. (Pat). (2015). Repetitive DNA in eukaryotic genomes. Chromosome Res. 23, 415–420. doi: 10.1007/s10577-015-9499-z

Brindley, P. J., Laha, T., McManus, D. P., and Loukas, A. (2003). Mobile genetic elements colonizing the genomes of metazoan parasites. Trends Parasitol. 19, 79–87. doi: 10.1016/S1471-4922(02)00061-2

Buonfrate, D., Formenti, F., Perandin, F., and Bisoffi, Z. (2015). Novel approaches to the diagnosis of Strongyloides stercoralis infection. Clin. Microbiol. Infect. 21, 543–552. doi: 10.1016/j.cmi.2015.04.001

Chanteau, S., Luquiaud, P., Failloux, A. B., and Williams, S. A. (1994). Detection of Wuchereria bancrofti larvae in pools of mosquitoes by the polymerase chain reaction. Trans. R. Soc. Trop. Med. Hyg. 88, 665–666. doi: 10.1016/0035-9203(94)90219-4

Charlesworth, B., Sniegowski, P., and Stephan, W. (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371, 215. doi: 10.1038/371215a0

Chidambaram, M., Parija, S. C., Toi, P. C., Mandal, J., Sankaramoorthy, D., George, S., et al. (2017). Evaluation of the utility of conventional polymerase chain reaction for detection and species differentiation in human hookworm infections. Trop. Parasitol. 7, 111. doi: 10.4103/tp.TP_26_17

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

Clarke, N. E., Llewellyn, S., Traub, R. J., McCarthy, J., Richardson, A., Nery, S. V. (2018). 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. 99, 1033–1040. doi: 10.4269/ajtmh.18-0356

Coghlan, A., Tyagi, R., Cotton, J. A., Holroyd, N., Rosa, B. A., Tsai, I. J., et al. (2019). Comparative genomics of the major parasitic worms. Nat. Genet. 51,163–174. doi: 10.1038/s41588-018-0262-1

Cringoli, G., Rinaldi, L., Maurelli, M. P., and Utzinger, J. (2010). FLOTAC: new multivalent techniques for qualitative and quantitative copromicroscopic diagnosis of parasites in animals and humans. Nat. Protoc. 5, 503–515. doi: 10.1038/nprot.2009.235

Cushion, M. T., Brubaker, R., Ash, L. R., and Orihel, T. C. (1990). Atlas of Human Parasitology. 3rd ed. Chicago: American Society of Clinical Pathologists. ISBN 0-89189-292-3. doi: 10.1111/j.1550-7408.1992.tb04458.x

Demas, A., Oberstaller, J., DeBarry, J., Lucchi, N. W., Srinivasamoorthy, G., Sumari, et al. (2011). Applied genomics: data mining reveals species-specific malaria diagnostic targets more sensitive than 18S rRNA. *J. Clin. Microbiol.* 49, 2411–2418. doi: 10.1128/JCM.02603-10

Easton, A. V., Oliveira, R. G., O'Connell, E.M., Kepha, S., Mwandawiro, C. S., Njenga, S. M., et al. (2016). 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* 9, 38. doi: 10.1186/s13071-016-1314-y

Easton, A. V., Oliveira, R. G., Walker, M., O'Connell, E. M., Njenga, S. M., Mwandawiro, C. S., et al. (2017). Sources of variability in the measurement of Ascaris lumbricoides infection intensity by Kato-Katz and qPCR. *Parasit. Vectors* 10, 256. doi: 10.1186/s13071-017-2164-y

Farrell, S. H., and Anderson, R. M. (2018). Helminth lifespan interacts with non-compliance in reducing the effectiveness of anthelmintic treatment. *Parasit. Vectors* 11, 66. doi: 10.1186/s13071-018-2670-6

Fischer, K., Gankpala, A., Gankpala, L., Bolay, F.K., Curtis, K.C., Weil, G.J., et al. (2018). Capillaria Ova and Diagnosis of Trichuris trichiura Infection in Humans by Kato-Katz Smear, Liberia. *Emerg. Infect. Dis.* 24, 1551. doi: 10.3201/eid2408.180184

Flavell, R. B., Bennett, M. D., Smith, J. B., and Smith, D. B. (1974). Genome size and the proportion of repeated nucleotide sequence DNA in plants. *Biochem. Genet.* 12, 257–269. doi: 10.1007/BF00485947

Funakoshi, K., Bagheri, M., Zhou, M., Suzuki, R., Abe, H., Akashi, H. (2017). Highly sensitive and specific Alu-based quantification of human cells among rodent cells. *Sci. Rep.* 7, 13202. doi: 10.1038/s41598-017-13402-3

Ganley, A. R. D., and Kobayashi, T. (2007). Highly efficient concerted evolution in the ribosomal DNA repeats: total rDNA repeat variation revealed by whole-genome shotgun sequence data. *Genome Res.* 17, 184–191. doi: 10.1101/gr.5457707

George, S., Kaliappan, S. P., Kattula, D., Roy, S., Geldhof, P., Kang, G., et al. (2015). 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.* 109, 283–285. doi: 10.1093/trstmh/trv001

Halfon, P., Sylla, K., Ndiaye, J. L., Delaunay, P., Gaye, O., Faye, B., et al. (2017). 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.* 97, 173–182. https://doi.org/10.4269/ajtmh.16-0781

Hamburger, J., He-Na, Abbasi, I., Ramzy, R. M., Jourdane, J., Ruppel, A. (2001). 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.* 65, 907–911.

Harmon, A. F., Zarlenga, D. S., and Hildreth, M. B. (2006). Improved methods for isolating DNA from Ostertagia ostertagi eggs in cattle feces. *Vet. Parasitol.* 135, 297–302. doi: 10.1016/j.vetpar.2005.10.014

Hicks, J. H., Kremer, M., and Miguel, E. (2015). The Case for Mass Treatment of Intestinal Helminths in Endemic Areas. *PLoS Negl. Trop. Dis.* 9, e0004214. doi: 10.1371/journal.pntd.0004214

Hidalgo, A., Melo, A., Romero, F., Hidalgo, V., Villanueva, J., Fonseca-Salamanca, F. (2018). DNA extraction in Echinococcus granulosus and Taenia spp. eggs in dogs stool samples applying thermal shock. *Exp. Parasitol.* 186, 10–16. doi: 10.1016/j.exppara.2018.01.016

Hii, S.F., Senevirathna, D., Llewellyn, S., Inpankaew, T., Odermatt, P., Khieu, V., et al. (2018). Development and Evaluation of a Multiplex Quantitative Real-Time Polymerase Chain Reaction for Hookworm Species in Human Stool. *Am. J. Trop. Med. Hyg.* 99, 1186–1193. doi: 10.4269/ajtmh.18-0276

Holt, D., Shield, J., Harris, T., Mounsey, K., Aland, K., McCarthy, J., et al. (2017). 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.* 2, 51. doi: 10.3390/tropicalmed2040051

Howe, K. L., Bolt, B. J., Cain, S., Chan, J., Chen, W. J., Davis, P., et al. (2016). WormBase 2016: expanding to enable helminth genomic research. *Nucleic Acids Res.* 44, D774–D780. doi: 10.1093/nar/gkv1217

Irisarri-Gutiérrez, M. J., Muñoz-Antolí, C., Acosta, L., Parker, L.A., Toledo, R., Bornay-Llinares, F.J., et al. (2016). Hookworm-like eggs in children's faecal samples from a rural area of Rwanda. *Afr. Health Sci.* 16, 83–88. https://doi.org/10.4314/ahs.v16i1.11

Kaisar, M.M.M., Brienen, E. a. T., Djuardi, Y., Sartono, E., Yazdanbakhsh, M., Verweij, J. J., et al. (2017). 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* 144, 965–974. doi: 10.1017/S0031182017000129

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

Knopp, S., Salim, N., Schindler, T., Voules, D. A. K., Rothen, J., Lweno, O., et al. (2014). 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.* 90, 535–545. doi: 10.4269/ajtmh.13-0268

Kongs, A., Marks, G., Verlé, P., and Stuyft, P. V. der (2001). The unreliability of the Kato-Katz technique limits its usefulness for evaluating *S. mansoni* infections. *Trop. Med. Int. Health* 6, 163–169. doi: 10.1046/j.1365-3156.2001.00687.x

Krauth, S. J., Coulibaly, J. T., Knopp, S., Traoré, M., N'Goran, E. K., Utzinger, J. (2012). An In-Depth Analysis of a Piece of Shit: Distribution of Schistosoma mansoni and Hookworm Eggs in Human Stool. *PLoS Negl. Trop. Dis.* 6, e1969. doi: 10.1371/journal.pntd.0001969

Kyu, H. H., Abate, D., Abate, K. H., Abay, S. M., Abbafati, C., Abbasi, N., et al. (2018). 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* 392, 1859–1922. doi: 10.1016/S0140-6736(18)32335-3

Lafontaine, D. L. J., and Tollervey, D. (2001). The function and synthesis of ribosomes. *Nat. Rev. Mol. Cell Biol.* 2, 514. doi: 10.1038/35080045

Lim, M. D., Brooker, S. J., Belizario, V. Y., Gay-Andrieu, F., Gilleard, J., Levecke, B., et al., Annecy STH diagnostic experts group (2018). Diagnostic tools for soil-transmitted helminths control and elimination programs: A pathway for diagnostic product development. *PLoS Negl. Trop. Dis.* 12, e0006213. doi: 10.1371/journal.pntd.0006213

Lindner, D. L., and Banik, M. T. (2011). Intragenomic variation in the ITS rDNA region obscures phylogenetic relationships and inflates estimates of operational taxonomic units in genus *Laetiporus. Mycologia* 103, 731–740. doi: 10.3852/10-331

Llewellyn, S., Inpankaew, T., Nery, S. V., Gray, D. J., Verweij, J. J., Clements, A. C. A., et al. (2016). Application of a Multiplex Quantitative PCR to Assess Prevalence and Intensity Of Intestinal Parasite Infections in a Controlled Clinical Trial. *PLoS Negl. Trop. Dis.* 10. doi: 10.1371/journal.pntd.0004380

Lodh, N., Caro, R., Sofer, S., Scott, A., Krolewiecki, A., Shiff, C. (2016). Diagnosis of Strongyloides stercoralis : Detection of parasite-derived DNA in urine. *Acta Trop.* 163, 9–13. doi: 10.1016/j.actatropica.2016.07.014

López-Flores, I., and Garrido-Ramos, M. A. (2012). "The repetitive DNA content of eukaryotic genomes," in *Genome Dynamics*, vol. 7. Eds. Garrido-Ramos and M. A. (S. KARGER AG), 1–28. doi: 10.1159/000337118

Mascagni, F., Vangelisti, A., Giordani, T., Cavallini, A., and Natali, L. (2018). Specific LTR-retrotransposons show copy number variations between wild and cultivated sunflowers. *Genes* 9(9), 433. doi: 10.3390/genes9090433

Masny, A., Sałamatin, R., Rozej-Bielicka, W., and Golab, E. (2016). Is molecular xenomonitoring of mosquitoes for *Dirofilaria repens* suitable for dirofilariosis surveillance in endemic regions? *Parasitol. Res.* 115, 511–525. doi: 10.1007/s00436-015-4767-6

Maurelli, M. P., Rinaldi, L., Alfano, S., Pepe, P., Coles, G. C., Cringoli, G. (2014). Mini-FLOTAC, a new tool for copromicroscopic diagnosis of common intestinal nematodes in dogs. *Parasit. Vectors* 7, 356. doi: 10.1186/1756-3305-7-356

McReynolds, L. A., DeSimone, S. M., and Williams, S. A. (1986). Cloning and comparison of repeated DNA sequences from the human filarial parasite *Brugia malayi* and the animal parasite *Brugia pahangi. Proc. Natl. Acad. Sci. U.S.A.* 83, 797–801. doi: 10.1073/pnas.83.3.797

Medley, G. F., Turner, H. C., Baggaley, R. F., Holland, C., and Hollingsworth, T. D. (2016). "Chapter Six—The Role of More Sensitive Helminth Diagnostics in Mass Drug Administration Campaigns: Elimination and Health Impacts," in *Advances in Parasitology*, vol. 94. Eds. Basáñez, M. G. , Anderson, and R. M. (Academic Press), 343–392. doi: 10.1016/bs.apar.2016.08.005

Meredith, S. E., Unnasch, T. R., Karam, M., Piessens, W. F., and Wirth, D. F. (1989). Cloning and characterization of an *Onchocerca volvulus* specific DNA sequence. *Mol. Biochem. Parasitol.* 36, 1–10. doi: 10.1016/0166-6851(89)90194-1

Mines, J. (1977). Modifications of Mcmaster worm egg counting method. *Aust. Vet. J.* 53, 342–343. doi: 10.1111/j.1751-0813.1977.tb00247.x

Mohd-Shaharuddin, N., Lim, Y. A. L., Hassan, N.-A., Nathan, S., and Ngui, R. (2019). Molecular characterization of *Trichuris* species isolated from humans, dogs and cats in a rural community in Peninsular Malaysia. *Acta Trop.* 190, 269–272. doi: 10.1016/j.actatropica.2018.11.026

Moser, W., Bärenbold, O., Mirams, G.J., Cools, P., Vlaminck, J., Ali, S.M., et al. (2018). Diagnostic comparison between FECPAKG2 and the Kato-Katz method for analyzing soil-transmitted helminth eggs in stool. *PLoS Negl. Trop. Dis.* 12, e0006562. doi: 10.1371/journal.pntd.0006562

Nekrutenko, A., Makova, K. D., and Baker, R. J. (2000). Isolation of binary species-specific PCR-based markers and their value for diagnostic applications. *Gene* 249, 47–51. doi: 10.1016/S0378-1119(00)00168-2

Nikolay, B., Brooker, S. J., and Pullan, R. L. (2014). Sensitivity of diagnostic tests for human soil-transmitted helminth infections: a meta-analysis in the absence of a true gold standard. *Int. J. Parasitol.* 44, 765–774. doi: 10.1016/j.ijpara.2014.05.009

Ng-Nguyen, D., Stevenson, M. A., Dorny, P., Gabriël, S., Vo, T. V., Nguyen, V.-A.T., et al. (2017). 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. 11, e0005743. doi:10.1371/journal.pntd.0005743

Novák, P., Neumann, P., and Macas, J. (2010). Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinformatics* 11, 378. doi: 10.1186/1471-2105-11-378

Novák, P., Neumann, P., Pech, J., Steinhaisl, J., and Macas, J. (2013). RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* 29, 792–793. doi: 10.1093/bioinformatics/btt054

Novák, P., Ávila Robledillo, L., Koblížková, A., Vrbová, I., Neumann, P., Macas, J. (2017). TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. *Nucleic Acids Res. 45,* e111. doi: 10.1093/nar/gkx257

Nunes, C.M., Lima, L.G.F., Manoel, C.S., Pereira, R.N., Nakano, M.M., and Garcia, J.F. (2006). Fecal specimens preparation methods for PCR diagnosis of human taeniosis. *Rev. Inst. Med. Trop. de São Paulo* 48(1), 45-47. doi: 10.1590/S0036-46652006000100010

O'Connell, E. M., et al. (2018). *Ancylostoma ceylanicum* Hookworm in Myanmar Refugees, Thailand, 2012-2015. *Emerg. Infect. Dis.* 24. doi: 10.3201/eid2408.180280

Okorie, P. N., and de Souza, D. K. (2016). Prospects, drawbacks and future needs of xenomonitoring for the endpoint evaluation of lymphatic filariasis elimination programs in Africa. *Trans. R. Soc. Trop. Med. Hyg.* 110, 90–97. doi: 10.1093/trstmh/trv104

Onyabe, D. Y., and Conn, J. E. (1999). Intragenomic heterogeneity of a ribosomal DNA spacer (ITS2) varies regionally in the neotropical malaria vector *Anopheles nuneztovari* (Diptera: Culicidae). *Insect Mol. Biol.* 8, 435–442. doi: 10.1046/j.1365-2583.1999.00134.x

Papaiakovou, M., Pilotte, N., Grant, J.R., Traub, R.J., Llewellyn, S., McCarthy, et al. (2017). 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.* 11, e0005734. doi: 10.1371/journal.pntd.0005734

Papaiakovou, M., Pilotte, N., Baumer, B., Grant, J., Asbjornsdottir, K., Schaer, F., et al. (2018). A comparative analysis of preservation techniques for the optimal molecular detection of hookworm DNA in a human fecal specimen. PLoS *Negl. Trop. Dis.* 12, e0006130. doi: 10.1371/journal.pntd.0006130

Pickering, A.J., Njenga, S.M., Steinbaum, L., Swarthout, J., Lin, A., Arnold, B.F., et al. (2019). 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.* 16, e1002841. doi: 10.1371/journal.pmed.1002841

Piednoël, M., Aberer, A.J., Schneeweiss, G.M., Macas, J., Novak, P., Gundlach, H., et al. (2012). Next-generation sequencing reveals the impact of repetitive DNA across phylogenetically closely related genomes of orobanchaceae. *Mol. Biol. Evol.* 29, 3601–3611. doi: 10.1093/molbev/mss168

Pilotte, N., Papaiakovou, M., Grant, J.R., Bierwert, L.A., Llewellyn, S., McCarthy, J.S., et al. (2016a). Improved PCR-based detection of soil transmitted helminth infections using a next-generation sequencing approach to assay design. *PLoS Negl. Trop. Dis.* 10, e0004578. doi: 10.1371/journal.pntd.0004578

Pilotte, N., Unnasch, T. R., and Williams, S. A. (2017). The current status of molecular xenomonitoring for lymphatic filariasis and onchocerciasis. *Trends Parasitol.* 33, 788–798. doi: 10.1016/j.pt.2017.06.008

Pilotte, N., Zaky, W. I., Abrams, B. P., Chadee, D. D., and Williams, S. A. (2016b). A novel xenomonitoring technique using mosquito excreta/feces for the detection of filarial parasites and malaria. *PLoS Negl. Trop. Dis.* 10, e0004641. doi: 10.1371/journal.pntd.0004641

Plohl, M., Luchetti, A., Mestrović, N., and Mantovani, B. (2008). Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. *Gene* 409, 72–82. doi: 10.1016/j.gene.2007.11.013

Pontes, L. A., Oliveira, M. C., Katz, N., Dias-Neto, E., and Rabello, A. (2003). Comparison of a polymerase chain reaction and the Kato-Katz technique for diagnosing infection with *Schistosoma mansoni. Am. J. Trop. Med. Hyg.* 68, 652–656. doi: 10.4269/ajtmh.2003.68.652

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., et al. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* 35, 7188–7196. doi: 10.1093/nar/gkm864

Ralph, A., O'Sullivan, M. V. N., Sangster, N. C., and Walker, J. C. (2006). Abdominal pain and eosinophilia in suburban goat keepers—trichostrongylosis. *Med. J. Aust.* 184, 467–469.

Ramírez, A.L., Hall-Mendelin, S., Doggett, S.L., Hewitson, G.R., McMahon, J.L., Ritchie, S.A., et al. (2018). Mosquito excreta: a sample type with many potential applications for the investigation of Ross River virus and West Nile virus ecology. *PLoS Negl. Trop. Dis.* 12, e0006771. doi: 10.1371/journal.pntd.0006771

Rao, R. U., Weil, G. J., Fischer, K., Supali, T., and Fischer, P. (2006). Detection of *Brugia* parasite DNA in human blood by real-time PCR. *J. Clin. Microbiol.* 44, 3887–3893. doi: 10.1128/JCM.00969-06

Rao, U. R., Williams, S. A., and Klei, T. R. (2002). Quantification of PCR amplification products of Brugia HhaI repeat DNA using a semiautomated Q-PCR system. *Mol. Cell. Probes* 16, 13–23. doi: 10.1006/mcpr.2001.0393

Sanpool, O., Intapan, P.M., Thanchomnang, T., Sri-Aroon, P., Lulitanond, V., Sadaow, L., et al. (2012). Development of a real-time PCR assay with fluorophore-labelled hybridization probes for detection of Schistosoma mekongi in infected snails and rat feces. *Parasitology* 139, 1266–1272. doi: 10.1017/S0031182012000649

Sarhan, R. M., Kamel, H. H., Saad, G. A., and Ahmed, O. A. (2015). Evaluation of three extraction methods for molecular detection of *Schistosoma mansoni* infection in human urine and serum samples. *J. Parasit. Dis. Off. Organ Indian Soc. Parasitol.* 39, 499–507. doi: 10.1007/s12639-013-0385-3

Schaap, M., Lemmers, R.J., Maassen, R., van der Vliet, P.J., Hoogerheide, L.F., van Dijk, H.K., et al. (2013). 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* 14, 143. doi: 10.1186/1471-2164-14-143

Shapiro, J. A., and Sternberg, R. (2005). von. Why repetitive DNA is essential to genome function. *Biol. Rev.* 80, 227–250. doi: 10.1017/S1464793104006657

Speich, B., Ali, S.M., Ame, S.M., Albonico, M., Utzinger, J., Keiser, J. (2015). Quality control in the diagnosis of Trichuris trichiura and Ascaris lumbricoides using the Kato-Katz technique: experience from three randomised controlled trials. *Parasit. Vectors* 8, 82. doi: 10.1186/s13071-015-0702-z

Stensvold, C.R., Saijuntha, W., Sithithaworn, P., Wongratanacheewin, S., Strandgaard, H., Ørnbjerg, N., et al. (2006). Evaluation of PCR based coprodiagnosis of human opisthorchiasis. *Acta Trop.* 97, 26–30. doi: 10.1016/j.actatropica.2005.08.008

Subirana, J. A., and Messeguer, X. (2013). A satellite explosion in the genome of holocentric nematodes. *PLoS One* 8, e62221. doi: 10.1371/journal.pone.0062221

Subirana, J. A., Albà, M. M., and Messeguer, X. (2015). High evolutionary turnover of satellite families in *Caenorhabditis. BMC Evol. Biol.* 15, 218. doi: 10.1186/s12862-015-0495-x

Teruel, M., Ruíz-Ruano, F.J., Marchal, J.A., Sánchez, A., Cabrero, J., Camacho, J.P., et al. (2014). Disparate molecular evolution of two types of repetitive DNAs in the genome of the grasshopper Eyprepocnemis plorans. *Heredity* 112, 531–542. doi: 10.1038/hdy.2013.135

Truscott, J. E., Turner, H. C., Farrell, S. H., and Anderson, R. M. (2016). Soil-transmitted helminths: mathematical models of transmission, the impact of mass drug administration and transmission elimination criteria. *Adv. Parasitol.* 94, 133–198. doi: 10.1016/bs.apar.2016.08.002

Truscott, J. E., Werkman, M., Wright, J.E., Farrell, S.H., Sarkar, R., Ásbjörnsdóttir, K., et al. (2017). Identifying optimal threshold statistics for elimination of hookworm using a stochastic simulation model. *Parasit. Vectors* 10, 321. doi: 10.1186/s13071-017-2256-8

Truscott, J. E., Dunn, J. C., Papaiakovou, M., Schaer, F., Werkman, M., Littlewood, D. T. J., et al. (2019). Calculating the prevalence of soil-transmitted helminth infection through pooling of stool samples: choosing and optimizing the pooling strategy. *PLoS Negl. Trop. Dis.* 13, e0007196. doi: 10.1371/journal.pntd.0007196

Umemori, J., Mori, A., Ichiyanagi, K., Uno, T., and Koide, T. (2013). Identification of both copy number variation-type and constant-type core elements in a large segmental duplication region of the mouse genome. *BMC Genomics* 14, 455. doi: 10.1186/1471-2164-14-455

Uniting to Combat NTDs. (2012) London declaration on neglected tropical diseases. Available at: https://unitingtocombatntds.org/london-declaration-neglected-tropical-diseases/ (Accessed April 9, 2019).

Verweij, J. J., and Stensvold, C. R. (2014). Molecular testing for clinical diagnosis and epidemiological investigations of intestinal parasitic infections. *Clin. Microbiol. Rev.* 27, 371–418. doi: 10.1128/CMR.00122-13

Weatherhead, J. E., Hotez, P. J., and Mejia, R. (2017). The global state of helminth control and elimination in children. *Pediatr. Clin.* 64, 867–877. doi: 10.1016/j.pcl.2017.03.005

Werneck, J. S., Carniato, T., Gabriel, A., Tufik, S., and Andrade, S. S. (2007). Mites in clinical stool specimens: potential misidentification as helminth eggs. *Trans. R. Soc. Trop. Med. Hyg.* 101, 1154–1156. doi: 10.1016/j.trstmh.2007.07.006

Wickstead, B., Ersfeld, K., and Gull, K. (2003). Repetitive elements in genomes of parasitic protozoa. *Microbiol. Mol. Biol. Rev.* 67, 360–375. doi: 10.1128/MMBR.67.3.360-375.2003

Williams, S. A., Lizotte-Waniewski, M. R., Foster, J., Guiliano, D., Daub, J., Scott, A.L., et al. (2000). The filarial genome project: analysis of the nuclear, mitochondrial and endosymbiont genomes of Brugia malayi. *Int. J. Parasitol.* 30, 411–419. doi: 10.1016/S0020-7519(00)00014-X

Yong, T.-S., Lee, J.-H., Sim, S., Lee, J., Min, D.-Y., Chai, J.-L., et al. (2007). Differential diagnosis of *Trichostrongylus* and hookworm eggs *via* PCR using ITS-1 sequence. *Korean J. Parasitol.* 45, 69–74. doi: 10.3347/kjp.2007.45.1.69

Zaky, W. I., Tomaino, F. R., Pilotte, N., Laney, S. J., and Williams, S. A. (2018). Backpack PCR: a point-of-collection diagnostic platform for the rapid detection of *Brugia* parasites in mosquitoes. *PLoS Negl. Trop. Dis.* 12, e0006962. doi: 10.1371/journal.pntd.0006962

Zhao, S., and Gibbons, J. G. (2018). A population genomic characterization of copy number variation in the opportunistic fungal pathogen *Aspergillus fumigatus*. *PLoS One* 13, e0201611 doi: 10.1371/journal.pone.0201611.

Profiling Transcriptional Regulation and Functional Roles of *Schistosoma mansoni* c-Jun N-Terminal Kinase

Sandra Grossi Gava[1], Naiara Clemente Tavares[1], Franco Harald Falcone[2,3], Guilherme Oliveira[4] and Marina Moraes Mourão[1]*

[1] *Laboratório de Helmintologia e Malacologia Médica, Instituto René Rachou, Fundação Oswaldo Cruz, Belo Horizonte, Brazil,* [2] *Allergy and Infectious Diseases Laboratory, Division of Molecular Therapeutics and Formulation, School of Pharmacy, University of Nottingham, Nottingham, United Kingdom,* [3] *Institute of Parasitology, BFS, Justus Liebig University, Giessen, Germany,* [4] *Environmental Genomics Group, Instituto Tecnológico Vale, Belém, Brazil*

Mitogen-activated protein kinases (MAPKs) play a regulatory role and influence various biological activities, such as cell proliferation, differentiation, and survival. Our group has demonstrated through functional studies that *Schistosoma mansoni* c-Jun N-terminal kinase (SmJNK) MAPK is involved in the parasite's development, reproduction, and survival. SmJNK can, therefore, be considered a potential target for the development of new drugs. Considering the importance of SmJNK in *S. mansoni* maturation, we aimed at understanding of SmJNK regulated signaling pathways in the parasite, correlating expression data with *S. mansoni* development. To better understand the role of SmJNK in *S. mansoni* intravertebrate host life stages, RNA interference knockdown was performed in adult worms and in schistosomula larval stage. SmJNK knocked-down in adult worms showed a decrease in oviposition and no significant alteration in their movement. RNASeq libraries of SmJNK knockdown schistosomula were sequenced. A total of 495 differentially expressed genes were observed in the SmJNK knockdown parasites, of which 373 were down-regulated and 122 up-regulated. Among the down-regulated genes, we found transcripts related to protein folding, purine nucleotide metabolism, the structural composition of ribosomes and cytoskeleton. Genes coding for proteins that bind to nucleic acids and proteins involved in the phagosome and spliceosome pathways were enriched. Additionally, we found that SmJNK and Smp38 MAPK signaling pathways converge regulating the expression of a large set of genes. *C. elegans* orthologous genes were enriched for genes related to sterility and oocyte maturation, corroborating the observed phenotype alteration. This work allowed an in-depth analysis of the SmJNK signaling pathway, elucidating gene targets of regulation and functional roles of this critical kinase for parasite maturation.

Keywords: *Schistosoma mansoni*, mitogen-activated protein kinases, c-Jun N-terminal kinase, RNA interference, signaling pathways, gene expression

INTRODUCTION

Schistosomiasis is one of the most common human parasitic diseases whose socioeconomic impact affects people in developing nations in the tropics and sub-tropics. Preventive chemotherapy (PC) with Praziquantel is required in 52 countries, where only in 2017, 220 million people needed PC for schistosomiasis; however, only 44.9% of this demand was covered (WHO, 2018). Despite the control efforts, schistosomiasis continues to be a major public health problem, principally due to schistosomiasis-related morbidity (French et al., 2018; Nascimento et al., 2019).

Schistosoma mansoni kinome containing 252 eukaryotic kinase proteins (ePKs) was first described by Andrade and co-authors (Andrade et al., 2011) based on an earlier version of *S. mansoni* genome (Berriman et al., 2009). Recently, 351 kinase genes with evidence of being transcribed in nearly all adult stages were described, of which 268 were PKs and an additional 83 were non-PKs (Grevelding et al., 2017). Although protein kinases have been recognized for years as suitable targets for drug development (Cohen, 2002; Cai et al., 2017), experimental functional evidence exists for only 40 *S. mansoni* proteins, showing that there is still need for further research.

The c-Jun N-terminal kinase signaling pathway is involved in the developmental regulation of various organisms. Its role has already been demonstrated in oocyte maturation and embryogenesis of *Xenopus laevis* (Bagowski et al., 2001) and in the spindle assembly during the mouse oocyte meiotic maturation (Huang et al., 2011). In *S. mansoni*, SmJNK (Smp_172240) was previously characterized, demonstrating its importance for the establishment of infection in the mammalian host (Andrade et al., 2014). SmJNK knockdown caused damage to the adult worms' tegument and impaired the maturation of vitelline organs. This resulted in lower oviposition and greater susceptibility to the host immune system. Furthermore, the SmJNK ortholog in *Schistosoma haematobium* was pointed as one of the prioritized druggable kinase targets due to its essentiality based on lethal gene knock-down or knock-out phenotypes in other organisms (Stroehlein et al., 2015).

Here, we sought to assess the role of SmJNK in adult worms *in vitro*, further advancing in unraveling the functions of this critical protein and putative druggable target. Also, we elucidated genes that are target of regulation by SmJNK knockdown, aiming to identify novel parasite-specific targets in addition to contributing to a better understanding of this signaling pathway. Additionally, we checked for a crosstalk between SmJNK pathway and the previously reported data from Smp38 (Avelar et al., 2019). Finally, we identified specific targets that correlate with previously observed phenotypes of developmental impairment (Andrade et al., 2014; Avelar et al., 2019) which is highly desirable for the design of new anti-schistosome drugs.

METHODS

Parasites

S. mansoni adult worms of LE strain were recovered from hamster periportal perfusion 40 days after cercariae percutaneous infection (Pellegrino and Siqueira, 1956). Schistosomula were obtained by mechanical transformation of cercariae as previously described (Milligan and Jolly, 2011). Cercariae were supplied by the Mollusk Room 'Lobato Paraense' of the René Rachou Institute–FIOCRUZ, where the parasite cycle is routinely maintained. The sporocysts were prepared following the protocol previously described (Mourão et al., 2009). This work was approved by the Oswaldo Cruz Foundation's Ethics Committee for Animal Use (CEUA) under number LW12/16, according to the Brazilian national guidelines set out in Law 11794/08.

DsRNA Synthesis

For the dsRNA synthesis, an SmJNK mRNA fragment corresponding to a region of approximately 570 bp, previously cloned into pGEM-T Easy vector, was amplified by PCR. Primers and cycling conditions used in this study were previously designed and established by Andrade and collaborators (Andrade et al., 2014). After amplification, PCR products were separated on 1% agarose gels, purified using QIAquick Gel Extraction Kit (Qiagen) and used as the template for dsRNA synthesis. DsRNA synthesis was performed using the T7 RiboMAX Express RNAi Systems kit (Promega) according to the supplier's protocol. DsRNA integrity and annealing were verified on 1% agarose electrophoresis.

SmJNK Knockdown in Adult Worms by RNA Interference

After perfusion, males and females adult worms were washed and separated manually. Then, eight males and eight females were placed separately in each well containing 100 µL of RPMI 1640 medium with 25 µg of dsRNA, in two technical replicates. The worms were electroporated with specific SmJNK dsRNA or unspecific GFP dsRNA into 4 mm cuvettes at 125 V for 20 ms and cultivated in 24-well plates with 1 mL RPMI 1640 medium supplemented with 10% heat-inactivated Fetal Bovine Serum and 2% Penicillin/Streptomycin. Unless stated otherwise, all culture reagents were from Gibco, Thermo Fisher Scientific. Worm motility was assessed using the WormAssay software (Marcellino et al., 2012) for 10 days, in which eight worms were cultured in 24-well tissue culture plates containing 1 mL of medium. Similarly, to count the number of eggs laid, eight worm pairs were electroporated and cultured in 6-well plates, the medium was changed daily, and the eggs counted. Also, on days 3, 5, 7, and 10 after electroporation, two worm pairs per day were removed and macerated with TRIzol Reagent (Thermo Fisher Scientific) for RNA extraction as described below. SmJNK expression in adult worms was assessed by quantitative real-time PCR (RT-qPCR). Experiments were performed in three biological replicates. Data were analyzed using unpaired t-test with Welch's correction.

SmJNK Knockdown in Schistosomula by RNA Interference

Schistosomula cultures (~500,000 worms/condition) were maintained in bottles with 10 mL of Glasgow Minimum Essential Medium (GMEM) (Sigma-Aldrich) supplemented

with 0.2 µM triiodothyronine (Sigma-Aldrich); 0.1% glucose; 0.1% lactalbumin (Sigma-Aldrich); 20 mM HEPES; 0.5% MEM vitamin solution (Gibco, Thermo Fisher Scientific); 5% Schneider's Insect Medium (Sigma-Aldrich); 0.5 µM Hypoxanthine (Sigma-Aldrich), 1 µM hydrocortisone (Sigma-Aldrich), 1% Penicillin/Streptomycin (Gibco, Thermo Fisher Scientific), and 2% heat-inactivated Fetal Bovine Serum (Gibco, Thermo Fisher Scientific). SmJNK dsRNA was added shortly after schistosomula transformation to a final concentration of 100 nM. Cultures were kept in a Bio-Oxygen Demand incubator (B.O.D.) at 37°C, 5% CO_2, and 95% humidity. Two independent biological replicates were performed. Data were analyzed using unpaired t-test with Welch's correction.

Total RNA Extraction

RNA extraction was performed three, five, seven, and ten days after adult worm electroporation with SmJNK dsRNA. For schistosomula, the RNA extraction was performed two days after exposure to SmJNK dsRNA, as previously shown, reduction in transcript levels was greater at this exposure time (Andrade et al., 2014). For RNA extraction, TRIzol Reagent (Invitrogen) was used associated with the RNeasy Mini Kit (Qiagen) according to the manufacturer's guidelines. Samples were treated with Turbo DNase (Ambion, Thermo Fisher Scientific) for removal of genomic DNA. RNAs were quantified using a Qubit Fluorometer 2.0 (Invitrogen) and, for RNASeq, RNA integrity was assessed using the Agilent RNA 6000 Pico kit and BioAnalyzer 2100 (Agilent Technologies).

Gene Expression Analysis by Quantitative Real-Time PCR (RT-qPCR)

The cDNA was synthesized using SuperScript™ III Reverse Transcriptase (Invitrogen, Thermo Fisher Scientific) or Illustra Ready-To-Go RT-PCR Beads (GE Healthcare). qPCR was performed using Power SYBR® Green Master mix (Applied Biosystems, Thermo Fisher Scientific) on an ABI 7500 RT-PCR system (Applied Biosystems, Thermo Fisher Scientific) to evaluate SmJNK knockdown. Primers and cycling conditions used in this study were previously designed and established by Andrade and collaborators (Andrade et al., 2014). Cytochrome C oxidase I gene (Smp_900000) was used to normalize RNA input from schistosomula (Andrade et al., 2014). FAD dependent oxidoreductase domain containing protein (SmFAD, Smp_089880) and actin-related protein 10 (Sm-arp 10, Smp_093230) genes were used to normalize RNA input from adult worms (**Supplementary Table S1**). In each plate, two internal controls were included to assess both reagent and genomic DNA contamination (RNA samples). Post-RNAi, SmJNK transcript levels were analyzed using the relative $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001) and expressed as the percentage of difference compared to the untreated or unspecific control.

To evaluate SmJNK expression among the different stages of *S. mansoni* we used qPCR absolute quantification using copy number standards, employing a curve containing five points of an SmJNK clone 10-fold dilution. The copy number of each dilution was calculated by the ratio of the clone molecular mass to the Avogadro constant as described (Lee et al., 2006), and the absolute copy number of the SmJNK transcript at each stage was estimated by interpolation of the sample Ct using a standard curve, and expressed as copy number per ng of total RNA.

High-Throughput mRNA Sequencing (RNASeq)

After confirming SmJNK knockdown, 2 µg of total RNA extracted from untreated and SmJNK knocked-down schistosomula, two days after dsRNA exposure, were subjected to mRNA purification, fragmentation, cDNA synthesis, and library construction. Next generation barcode multiplexed sequencing libraries were constructed using the TruSeq stranded mRNA kit (Illumina) and pooled in equimolar concentrations. Libraries were sequenced as 100-base paired-end reads using HiSeq 2500 (Illumina) using a HiSeq Rapid SBS sequencing kit v.2 (Illumina). Raw sequence data were preprocessed using the standard Illumina pipeline to segregate multiplexed reads. Sequence quality was assessed using the FastQC program (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Reads were then mapped to the *S. mansoni* reference genome (v. 7) using the STAR program (v. 2.6.1d) using default parameters except the following: –outSJfilterReads Unique, –sjdbOverhang 99, –outFilterType BySJout, –outFilterMultimapNmax 20, –alignSJoverhangMin 8, –alignIntronMin 25, and –outSAMattributes All (Dobin et al., 2013).

Screening of Differentially Expressed Genes

Downstream analysis was carried out using the R statistical software package (v. 3.3.2) (R Development Core Team, 2011). To ascertain genes that significantly change after SmJNK dsRNA exposure, we used DESeq2 (v. 1.12.4) (Love et al., 2014) to normalize the data and performed statistical analysis. Genes with adjusted p-value <0.01 were considered significantly differentially expressed. Clustering analysis was performed using the heatmap function from the gplots (v. 3.0.1) package and PCA plots were generated using the ggplot2 (v. 2.1.1) package.

Gene ontology (GO) (Ashburner et al., 2000; The Gene Ontology Consortium, 2017) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000; Kanehisa et al., 2016; Kanehisa et al., 2017) pathway analysis were used to identify the enriched molecular functions and associated biological pathways of differentially expressed genes (DEGs). The g:Profiler tool (version e95_eg42_p13_f6e58b9) (Reimand et al., 2007; Raudvere et al., 2019) was applied to perform the GO enrichment and KEGG pathway analysis. The g:SCS method was applied for computing multiple testing correction using the padjusted values with a 0.05 threshold. The ReViGO tool (Supek et al., 2011) was used to cluster, summarize, and visualize the list of significantly enriched GO terms based on their semantic similarities (Allowed similarity: 0.4).

We also compared DEGs of schistosomula SmJNK knocked-down with those found in schistosomula Smp38 knocked-down (Avelar et al., 2019). The down-regulated DEGs that overlap in both datasets were subjected to the BioMart tool (https://parasite.

wormbase.org/info/Tools/biomart.html) from WormBase ParaSite (Howe et al., 2017) to search for orthologs in the nematode Caenorhabditis elegans. The orthologous genes were then analyzed using the Worm Enrichr enrichment analysis tool for *C. elegans* (https://amp.pharm.mssm.edu/WormEnrichr/) to search for the RNAi phenotypes associated with this gene list (Chen et al., 2013; Kuleshov et al., 2016).

Quantitative Real-Time PCR Validation of Differentially Expressed Genes

Quantitative PCR was performed to validate the expression levels of selected differentially expressed genes. Ten genes were selected randomly; whereas five of them are among the DEGs exhibiting large fold changes (**Supplementary Table S1**). Specific primers were designed using PrimerQuest tool (https://www.idtdna.com/Primerquest/Home/Index) and/or Primer3 software (http://primer3.ut.ee/) and obtained from Integrated DNA Technologies (IDT). The sequences of all primers and the final concentrations established for each qPCR reaction are presented in **Supplementary Table S1**. PCR efficiency for each pair of specific primers was estimated by titration analysis to be $100\% \pm 10\%$. RNA extraction, purification, cDNA synthesis, and RT-qPCR reactions were carried as described above. SmFAD and Sm-arp 10 genes were used to normalize RNA input (**Supplementary Table S1**).

RESULTS

SmJNK Expression Levels Among *S. Mansoni* Developmental Stages

The expression profile of SmJNK in developmental stages of *S. mansoni* (cercariae, two and seven days schistosomula, adult males, adult females, and sporocysts) was investigated by quantitative PCR. Absolute quantification was employed to assess
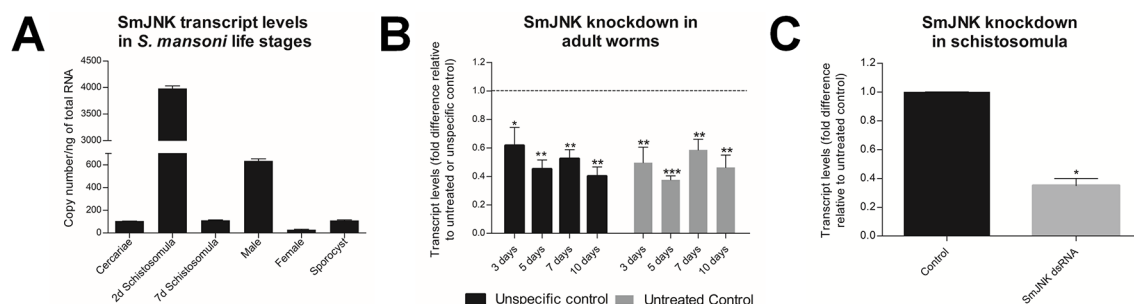
SmJNK expression among the different developmental stages. The SmJNK gene exhibited the highest expression levels (~4,000 copies/ng of total RNA) in two days schistosomula. This was six times the amount presented in adult males that exhibited the second highest expression levels (~635 copies/ng of total RNA) (**Figure 1A**). Cercariae, sporocysts and seven days transformed schistosomula presented approximately the same amount of SmJNK transcripts.

SmJNK Knockdown in Adult Schistosomes
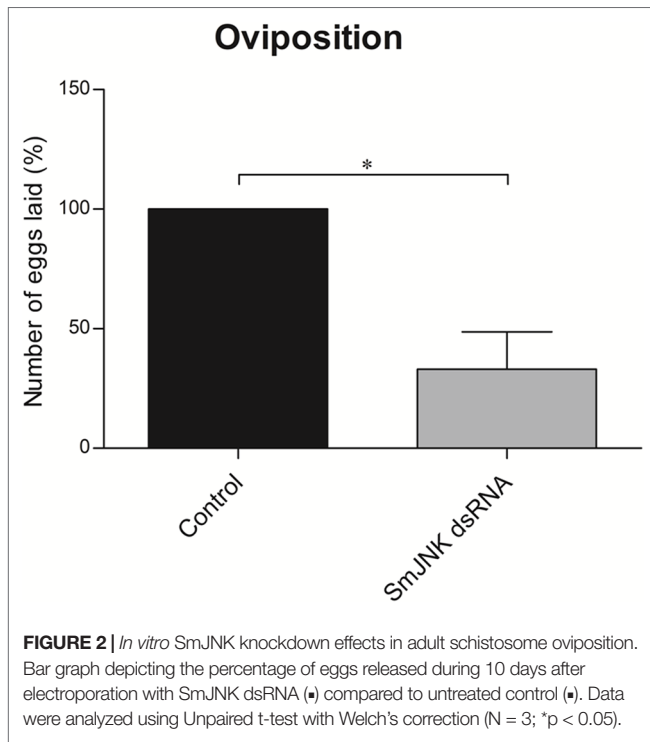
Adult worms electroporated with SmJNK dsRNA showed up to 62% reduction in SmJNK transcript levels on the fifth day after electroporation (**Figure 1B**). After the successful establishment of SmJNK knockdown in adults, we evaluated phenotypic changes, including worm movement and oviposition. SmJNK knockdown in adult worms did not result in any significant changes in male nor female movement for ten days (**Supplementary Figures S1A, B**). However, a significant reduction of 67% in egg laying was observed in worms electroporated with SmJNK dsRNA (**Figure 2**).

SmJNK Knockdown in Schistosomula and Quality Assessment of RNASeq Libraries

After confirming successful SmJNK knockdown by RT-qPCR in schistosomula (**Figure 1C**), four paired-end libraries were generated yielding a total of 158 M paired-end reads with 100 bps as average sequence length, with 38–40% guanine-cytosine (GC) content (**Table 1**). FastQC quality assessment indicated that RNASeq data presented a high quality and were adequate for downstream transcriptome analysis. Raw reads were submitted to Sequence Read Archive (SRA; http://www.ncbi.nlm.nih.gov/sra), under accession numbers: PRJNA354932 and PRJNA492452. Sequence mapping to *S. mansoni* reference

FIGURE 1 | SmJNK expression profile among developmental stages of *S. mansoni* and after RNAi of adult worms *in vitro*. **(A)** SmJNK transcript levels in the different *S. mansoni* life stages; cercariae, two and seven days schistosomula, males, females, and sporocysts. Bar graph depicting the absolute SmJNK transcript levels presented as copy number per ng of total RNA. Bars represent the standard deviation of the mean of three technical replicates. **(B)** Bar graph depicting the relative SmJNK transcript levels in adult worms three, five, seven, and ten days after electroporation with SmJNK dsRNA. Bars represent the relative values of SmJNK transcripts compared to the untreated control (■), or to the unspecific control-GFP (■). The dashed line represents the values of the normalized controls. Data are represented as mean fold-difference (+/− SE). Asterisks represent statistically significant differences. Data were analyzed using Unpaired t-test with Welch's correction (N=3, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$). **(C)** SmJNK transcript levels in schistosomula exposed to SmJNK dsRNA before RNASeq experiments. Bar graph depicting relative SmJNK transcripts levels in schistosomula two days after exposure to SmJNK dsRNA (■) compared to the untreated control (■). Data are represented as mean fold-differences (+/− SE). Transcript levels were determined by RT-qPCR and data analyzed using the ΔΔCt method (*$p < 0.05$, N = 2). Unpaired *t* test with Welch's correction.

FIGURE 2 | *In vitro* SmJNK knockdown effects in adult schistosome oviposition. Bar graph depicting the percentage of eggs released during 10 days after electroporation with SmJNK dsRNA (■) compared to untreated control (■). Data were analyzed using Unpaired t-test with Welch's correction (N = 3; *p < 0.05).

TABLE 1 | Summary of sequenced datasets and mapping to the *S. mansoni* reference genome.

Sample	Replicate	Input reads number	Uniquely mapped reads (%)	Reads mapped to multiple loci (%)
Control	1	46650550	80,94	15,35
	2	58555610	82,41	13,86
SmJNK	1	29013869	81,31	11,75
dsRNA	2	23856371	84,16	10,87
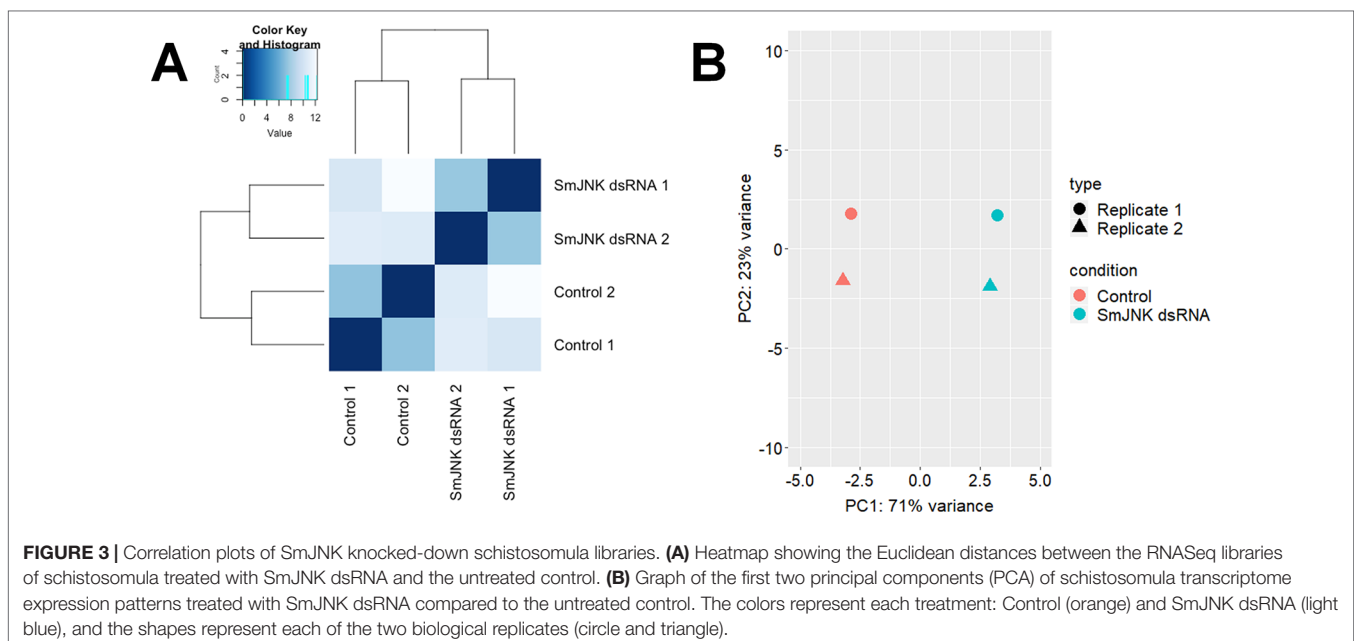
genome (v. 7) resulted in uniquely mapped reads ranging from 80.94 - 84.16%.

The correlation degree between the biological samples was investigated by hierarchical clustering using the heatmap function of the DESeq2 package and means of the Euclidean distances. Clustering analysis grouped libraries from biological replicates in the same branch (**Figure 3A**). Principal component analysis (PCA) was also performed with the genes by multidimensional scaling of the data matrix (**Figure 3B**). The PCA plot of the first two components showed a clear separation between the control and SmJNK dsRNA treated samples in the first dimension.
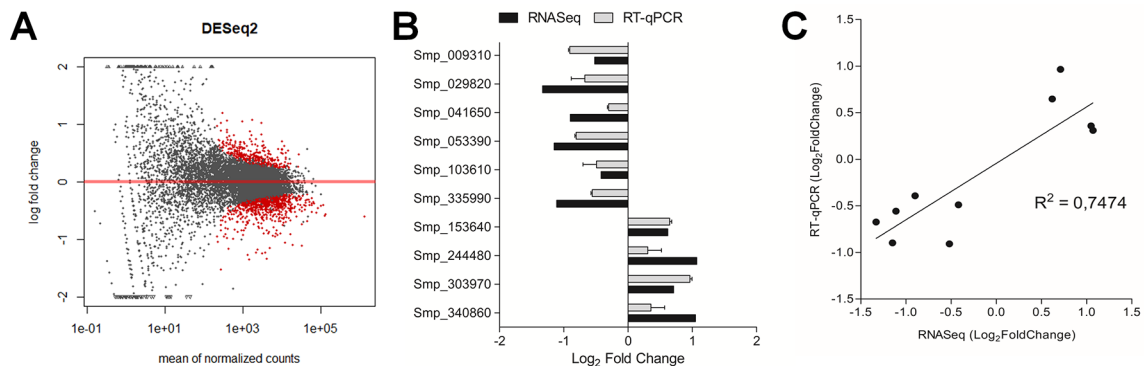
Potential SmJNK Regulated Gene Targets Identified in Schistosomula After RNAi

After checking for data consistency and sample quality, the transcriptional profile of schistosomula two days after SmJNK dsRNA exposure was compared to the profile of the untreated control using the DESeq2 package in R. We found 495 DEGs using the DESeq2 package (padj <0.01) (**Figure 4A**). Ten of the DEGs identified among the transcripts were selected for validation by RT-qPCR (**Figure 4B**) (**Supplementary Table S1**). All the analyzed genes showed consistent expression trend between the RNASeq and RT-qPCR analyses, although with some differences in $\log_2$FoldChange values, which is expected between two different methodologies and among different biological replicates. Significant correlation ($R^2 = 0.7474$, $p = 0.0012$) between the $\log_2$FoldChange of the employed methodologies confirms the accuracy of the data (**Figure 4C**).

A list of all statistically significant DEGs from schistosomula exposed to SmJNK dsRNA can be found in the **Supplementary Table S2**. Of the 495 DEGs detected in the samples in which SmJNK expression was reduced, 373 are down-regulated and 122 up-regulated in comparison to untreated controls.

FIGURE 3 | Correlation plots of SmJNK knocked-down schistosomula libraries. **(A)** Heatmap showing the Euclidean distances between the RNASeq libraries of schistosomula treated with SmJNK dsRNA and the untreated control. **(B)** Graph of the first two principal components (PCA) of schistosomula transcriptome expression patterns treated with SmJNK dsRNA compared to the untreated control. The colors represent each treatment: Control (orange) and SmJNK dsRNA (light blue), and the shapes represent each of the two biological replicates (circle and triangle).

FIGURE 4 | Differentially expressed genes (DEGs) after SmJNK knockdown in schistosomula. **(A)** MA plot depicting a DESeq2 analysis to identify DEGs between SmJNK dsRNA and untreated control. The log-fold change for each transcript is plotted against the mean of normalized counts; each point corresponds to one gene. Significantly altered gene expressions are highlighted in red (padj < 0.01). **(B)** RT-qPCR validation of the differentially expressed genes in response to SmJNK knockdown. Ten DEGs were selected from a range of up-regulated and down-regulated genes. Expression levels were quantified by RT-qPCR (gray) and the results were compared to those obtained by the RNASeq approach (black). **(C)** Pearson's correlation of Log$_2$FoldChange in differentially expressed transcripts between RNASeq and RT-qPCR analysis.

Gene Set Enrichment Analysis

The g:Profiler tool was used to elucidate the functional roles of the 495 DEGs. DEGs were categorized into three GO categories (GO Biological Process, GO Molecular Function, and GO Cell Component) and KEGG Pathways. The significantly enriched subcategories are shown in **Figure 5** and **Supplementary Table S3**. We observed a decrease in expression for genes encoding proteins related to (1) the ribosome structural composition, (2) cytoskeleton-related, (3) purine nucleotide metabolism (4) protein folding, (5) splicing mechanisms, (6) phagosomes, and (7) binding to nucleic acids. For DEGs that showed increased expression, no subcategory showed significant enrichment (**Figure 5A**).

FIGURE 5 | Gene Ontology Enrichment of down-regulated genes from SmJNK knockdown schistosomula. **(A)** Manhattan plot illustrating the enrichment analysis results separated into four categories: GO : MF (Molecular Function), GO : BP (Biological Process), GO : CC (Cellular Component), and KEGG Pathways. The number in the source name in the x-axis labels shows how many significantly enriched terms were found. The circle corresponds to term size. The term location on the x-axis is fixed and terms from the same GO subtree are located closer to each other. Interactive graph of the representative subset of the GO terms related to biological process **(B)** and molecular function **(C)** enriched in down-regulated genes from SmJNK knockdown schistosomula. Bubble colors indicate the p-value and size indicates the frequency of the GO term in the underlying Gene Ontology Annotation database. Highly similar GO terms are linked by edges, where the line width indicates the degree of similarity.

After summarizing the list of GO terms using the ReViGO tool, we found not redundant GO terms and ordered according to the $\log_{10}$ p-value. Focusing on enriched GO categories from Biological Process categories found among the down-regulated genes, we found that 23 GOs were representative in SmJNK knockdown schistosomula, due to their uniqueness, we highlight: protein folding, biosynthetic process, adenosine triphosphate (ATP) hydrolysis coupled cation transmembrane transport, ATP metabolic process, regulation of cellular catabolic process, gene expression, nitrogen compound metabolic process, RNA splicing, cellular amide metabolic process, protein import into nucleus, positive regulation of proteolysis, translational initiation, and cellular nitrogen compound metabolic process (**Figure 5B**, **Table 2**). The GO terms included in the molecular function category were summarized in 22 GO terms. We draw attention to structural constituent of ribosome, hydrogen ion transmembrane transporter activity, unfolded protein binding, lactate dehydrogenase activity, peptidyl-prolyl cis-trans isomerase activity, drug binding, translation initiation factor activity, and nucleic acid binding (**Figure 5C**, **Table 2**).

Crosstalk Between SmJNK and Smp38 MAPK Pathways

Since the JNK and p38 MAPK signaling pathways are triggered by multiple stimuli that simultaneously activate both pathways and consequently share several upstream regulators (Wagner and Nebreda, 2009), we compared the DEGs found here with those found for Smp38 knocked-down schistosomula (Avelar et al., 2019). The SmJNK and Smp38 MAPK pathways in *S. mansoni* presented an intersection of 311 and 89 genes with decreased and increased expression, respectively (**Figure 6**).

Of the 311 down-regulated genes identified in common for SmJNK and Smp38 knockdown schistosomula, we found orthologs for 221 of them, which correspond to 291 different genes in *C. elegans*, since some of the genes present more than one ortholog. The *C. elegans* orthologs are enriched in genes related to the RNAi phenotypes as described: sterile, embryonic lethal, maternal sterile, larval arrest, transgene subcellular localization variant, slow growth, among others (**Figure 7** and **Supplementary Table S4**).

DISCUSSION

We advanced the understanding of SmJNK gene function by demonstrating that SmJNK is involved in parasite reproduction *in vitro*. We observed a significantly lower number of eggs recovered from SmJNK knocked-down adult worms compared to the control group. It was previously shown that SmJNK knockdown decreased parasite survival and maturation within the definitive host, and that female worms treated with SmJNK dsRNA showed undifferentiated oocytes (Andrade et al., 2014). SmJNK is highly expressed in the ovary of paired females (Lu et al., 2016; Lu et al., 2017), evidencing its role in female reproduction. Therefore, in *in vitro* adult females, SmJNK knockdown may influence oocyte maturation, since the ovary is already fully formed prior to the SmJNK knockdown. This is particularly promising as the eggs

are fundamentally related to the disease pathology. Granulomas are the result of host immune response to schistosome eggs. Its formation is triggered by immune cells migration to encapsulate the eggs, resulting in damage of liver tissue (fibrosis and, ultimately, cirrhosis) (Espírito Santo et al., 2008; King and Dangerfield-Cha, 2008; Burke et al., 2009). Additionally, in *in vitro* cultured adult worms, SmJNK does not seem to have a role in parasite movement or viability.

Protein kinases domains are highly conserved between organisms (Manning et al., 2002). Therefore, any drug development effort needs to consider cross-binding of a small molecules that could result in undesired potential side effects. Inhibitors need to be specific and to target essential proteins. The elucidation of gene targets regulated by SmJNK, in addition to contributing to a better understanding of the JNK signaling pathway in *S. mansoni* parasite, also aims at searching for parasite-specific targets. Among the identified DEGs are some proteins with unknown function, these genes could be schistosome-specific. Such proteins may be related to the previously observed phenotypes of parasite development and oviposition impairment.

We observed the down-regulation of genes related to the structural composition of ribosomes in SmJNK knockdown parasites. Ribosomal proteins are major components of ribosomes and play critical roles in protein biosynthesis. Variations observed in the expression of ribosomal proteins in different human tissues are probably related to extraribosomal functions (Bortoluzzi et al., 2001), since ribosomal proteins and RNAs are typically synthesized in stoichiometric amounts (Mager, 1988). Also, the differential expression of ribosomal proteins has been reported in several pathological (Bévort and Leffers, 2000) and stress conditions (Wang et al., 2013).

Genes related to the cytoskeleton showed decreased expression when SmJNK was knocked-down. Integrin-linked kinase (ILK) an activator of the SmJNK, extracellular signal-regulated kinase (SmERK), and Akt pathways, is involved in cytoskeletal reorganization and cell survival. The dysregulation of SmILK may contribute to errors in cell division and genomic instability (Fielding and Dedhar, 2009). The SmILK signaling pathway has an influence on egg production, ovarian structure, and oocyte integrity in female schistosomes (Gelmedin et al., 2017), with similar phenotypes to those observed for SmJNK knockdown (Andrade et al., 2014).

Among genes with decreased expression after SmJNK knockdown, we observed genes related to protein folding, such as chaperonins and heat shock proteins (HSPs). Cellular chaperones such as HSPs confer resistance to stress, promoting cell survival. Prolonged oxidative stress causes an increase in misfolded proteins that aggregate in the endoplasmic reticulum. The objective of the endoplasmic reticulum stress response is to inhibit or retard protein synthesis. This is achieved by the phosphorylation of eukaryotic translation initiation factor 2 (eIF2a) by protein kinase R–like endoplasmic reticulum kinase or by JNK (Nagiah et al., 2016). SmJNK knockdown may have inhibited this stress response, resulting in the observed down-regulation of chaperone and HSPs gene expression.

In rat fibroblasts, positive regulation of the p38 and JNK signaling pathways promotes increased expression of the

TABLE 2 | List of enriched gene ontology (GO) from DEGs down-regulated in schistosomula SmJNK knockdown. The analysis is separated into two GO categories: Biological Process and Molecular Function. The dispensable GO terms are depicted in italic. The values of $\log_{10}$p-value refers to $\log_{10}$ of padjusted values obtained using g:Profiler toll.

	Biological process				
term ID	description	frequency	$\log_{10}$ p-value	uniqueness	dispensability
GO:0006457	protein folding	0.903%	−72.699	0.93	0.00
GO:0008150	biological_process	100,000%	−102.197	1.00	0.00
GO:0008152	metabolic process	75,387%	−79.533	0.99	0.00
GO:0009058	biosynthetic process	31,611%	−117.905	0.93	0.00
GO:0009987	cellular process	63,780%	−117.165	0.99	0.00
GO:0099132	ATP hydrolysis coupled cation transmembrane transport	0.729%	−46.910	0.79	0.00
GO:0015991	ATP hydrolysis coupled proton transport	0.163%	−14.544	0.77	0.80
GO:0099131	ATP hydrolysis coupled ion transmembrane transport	0.996%	−14.544	0.79	0.96
GO:0009206	purine ribonucleoside triphosphate biosynthetic process	0.485%	−25.175	0.56	0.98
GO:0009145	purine nucleoside triphosphate biosynthetic process	0.485%	−25.175	0.57	0.92
GO:0098662	inorganic cation transmembrane transport	1,858%	−13.551	0.79	0.86
GO:0090662	ATP hydrolysis coupled transmembrane transport	1,001%	−14.544	0.85	0.46
GO:1902600	hydrogen ion transmembrane transport	1,015%	−42.701	0.75	0.82
GO:0006754	ATP biosynthetic process	0.432%	−23.137	0.54	0.97
GO:0015988	energy coupled proton transmembrane transport, against electrochemical gradient	0.170%	−14.544	0.78	0.69
GO:0015985	energy coupled proton transport, down electrochemical gradient	0.411%	−23.137	0.76	0.88
GO:0015986	ATP synthesis coupled proton transport	0.411%	−23.137	0.42	0.88
GO:0009201	ribonucleoside triphosphate biosynthetic process	0.577%	−22.206	0.56	0.95
GO:0046034	ATP metabolic process	1,263%	−37.694	0.54	0.04
GO:0009205	purine ribonucleoside triphosphate metabolic process	1,366%	−25.175	0.57	0.97
GO:0009259	ribonucleotide metabolic process	2,752%	−18.803	0.58	0.92
GO:0019693	ribose phosphate metabolic process	3,032%	−18.803	0.66	0.67
GO:0006163	purine nucleotide metabolic process	2,435%	−18.803	0.55	0.94
GO:0009144	purine nucleoside triphosphate metabolic process	1,401%	−25.175	0.57	0.94
GO:0009142	nucleoside triphosphate biosynthetic process	0.651%	−22.206	0.57	0.86
GO:0009141	nucleoside triphosphate metabolic process	1,605%	−19.590	0.61	0.71
GO:0009150	purine ribonucleotide metabolic process	2,353%	−20.776	0.55	0.89
GO:0009199	ribonucleoside triphosphate metabolic process	1,458%	−22.206	0.57	0.96
GO:0031329	regulation of cellular catabolic process	0.093%	−22.161	0.66	0.05
GO:0009896	positive regulation of catabolic process	0.072%	−20.507	0.69	0.92
GO:1901575	organic substance catabolic process	4,612%	−14.509	0.71	0.60
GO:0009894	regulation of catabolic process	0.146%	−16.724	0.79	0.43
GO:1903364	positive regulation of cellular protein catabolic process	0.038%	−20.507	0.60	0.94
GO:1903362	regulation of cellular protein catabolic process	0.062%	−17.608	0.61	0.91
GO:1903052	positive regulation of proteolysis involved in cellular protein catabolic process	0.034%	−20.507	0.61	0.97
GO:1903050	regulation of proteolysis involved in cellular protein catabolic process	0.056%	−17.608	0.60	0.97
GO:0031331	positive regulation of cellular catabolic process	0.058%	−20.507	0.64	0.97
GO:0051603	proteolysis involved in cellular protein catabolic process	0.759%	−15.585	0.62	0.73
GO:0044257	cellular protein catabolic process	0.772%	−14.580	0.63	0.94
GO:1901800	positive regulation of proteasomal protein catabolic process	0.030%	−20.507	0.61	0.92
GO:0061136	regulation of proteasomal protein catabolic process	0.048%	−17.608	0.60	0.95
GO:0045732	positive regulation of protein catabolic process	0.054%	−20.507	0.62	0.90
GO:0010467	gene expression	19,671%	−154.664	0.81	0.07
GO:0044260	cellular macromolecule metabolic process	34276%	−79.682	0.75	0.42
GO:0044237	cellular metabolic process	53,061%	−116.849	0.87	0.08
GO:0006807	nitrogen compound metabolic process	38,744%	−86.404	0.93	0.09
GO:0044238	primary metabolic process	53,743%	−90.300	0.92	0.11
GO:0071704	organic substance metabolic process	58,357%	−83.098	0.92	0.12
GO:0043170	macromolecule metabolic process	39,491%	−78.359	0.85	0.18
GO:0008380	RNA splicing	0.413%	−16.333	0.77	0.20
GO:0043603	cellular amide metabolic process	6,879%	−235.632	0.78	0.20
GO:0006606	protein import into nucleus	0.102%	−15.977	0.82	0.24
GO:0034504	protein localization to nucleus	0.129%	−15.977	0.88	0.75
GO:0051170	nuclear import	0.107%	−13.547	0.87	0.88
GO:0045862	positive regulation of proteolysis	0.078%	−20.507	0.69	0.27

(Continued)

TABLE 2 | Continued
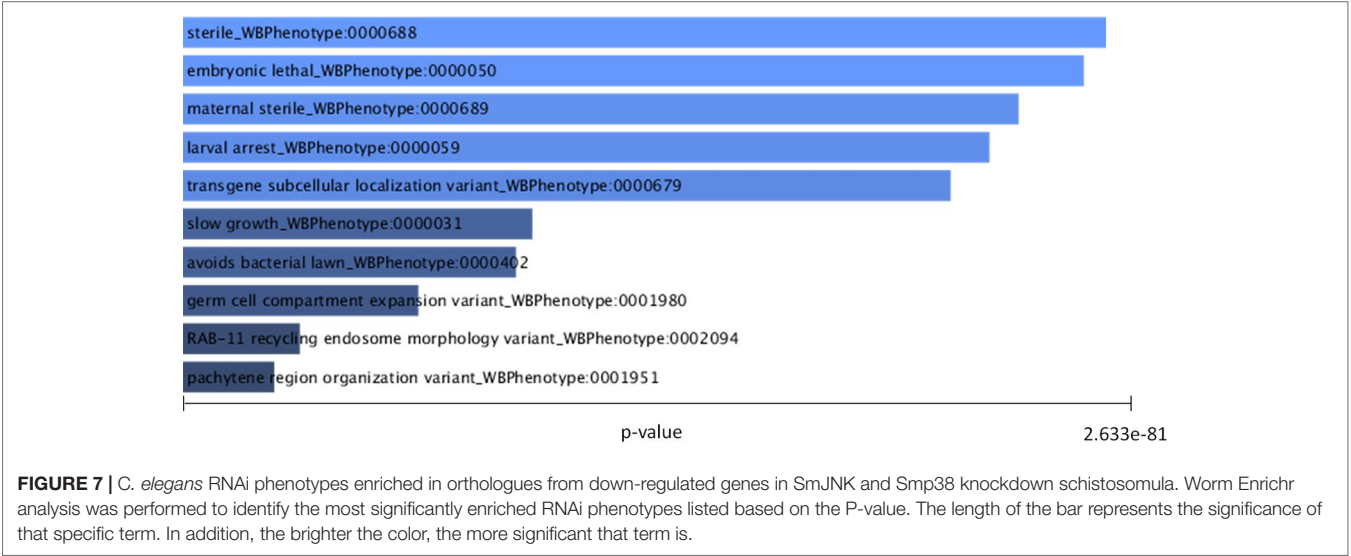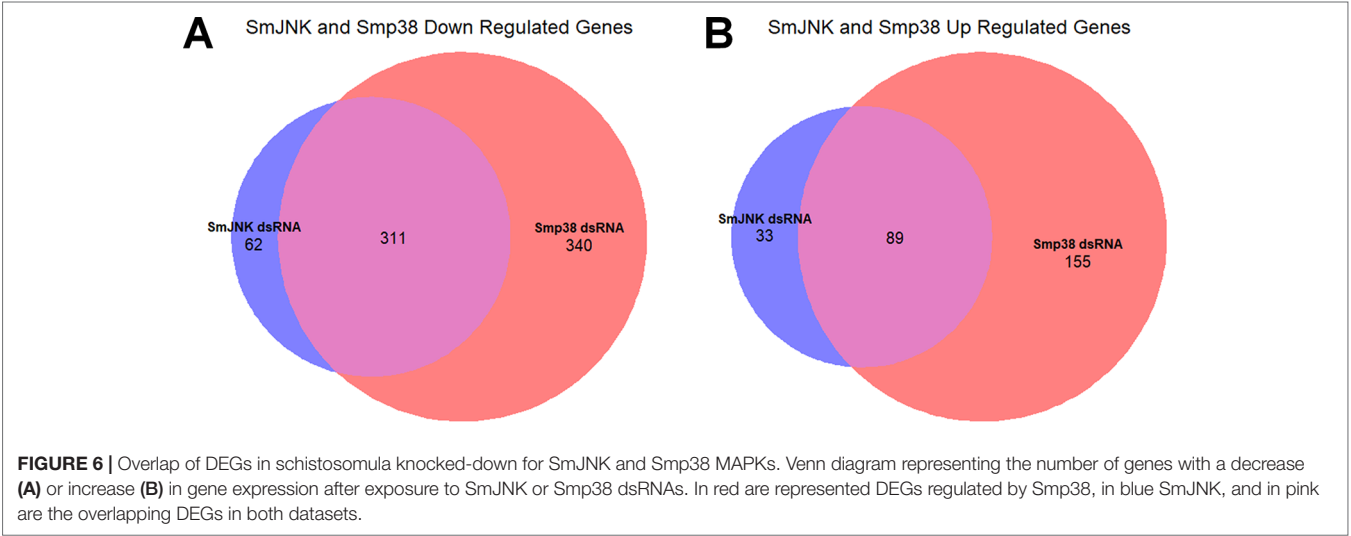
	Biological process				
term ID	description	frequency	log$_{10}$ p-value	uniqueness	dispensability
GO:0006413	translational initiation	0.518%	−13.840	0.65	0.29
GO:0006412	translation	5,686%	−258.077	0.55	0.66
GO:0006518	peptide metabolic process	5,961%	−248.427	0.67	0.88
GO:0043604	amide biosynthetic process	6,374%	−247.598	0.67	0.89
GO:0043043	peptide biosynthetic process	5,770%	−255.362	0.64	0.93
GO:0034641	cellular nitrogen compound metabolic process	34,137%	−138.052	0.74	0.31
GO:1901576	organic substance biosynthetic process	30,365%	−117.012	0.74	0.59
GO:1901564	organonitrogen compound metabolic process	17,886%	−107.455	0.77	0.42
GO:0009059	macromolecule biosynthetic process	19,548%	−131.642	0.72	0.50
GO:0044249	cellular biosynthetic process	30,048%	−115.243	0.73	0.66
GO:0044271	cellular nitrogen compound biosynthetic process	22,502%	−141.400	0.68	0.45
GO:0034645	cellular macromolecule biosynthetic process	19,291%	−133.236	0.68	0.54
GO:0019538	protein metabolic process	18,489%	−96.161	0.80	0.34
GO:0072521	purine-containing compound metabolic process	2,673%	−17.982	0.73	0.35
GO:1901566	organonitrogen compound biosynthetic process	14,064%	−207.359	0.66	0.51
GO:0044267	cellular protein metabolic process	14,293%	−109.414	0.72	0.38
	Molecular Function				
GO:0003674	molecular_function	100,000%	−174.435	1.00	0.00
GO:0003735	structural constituent of ribosome	2,679%	−276.945	0.90	0.00
GO:0005201	extracellular matrix structural constituent	0.028%	−52.190	0.91	0.56
GO:0005200	structural constituent of cytoskeleton	0.079%	−23.341	0.90	0.61
GO:0005198	structural molecule activity	3,268%	−331.562	0.95	0.00
GO:0005488	binding	55,656%	−113.354	0.98	0.00
GO:0015078	hydrogen ion transmembrane transporter activity	0.926%	−42.870	0.80	0.00
GO:0022853	active ion transmembrane transporter activity	0.923%	−17.206	0.80	0.68
GO:0044769	ATPase activity, coupled to transmembrane movement of ions, rotational mechanism	0.363%	−14.218	0.63	0.89
GO:0046933	proton-transporting ATP synthase activity, rotational mechanism	0.325%	−23.533	0.62	0.81
GO:0019829	cation-transporting ATPase activity	0.609%	−17.206	0.61	0.94
GO:0042625	ATPase coupled ion transmembrane transporter activity	0.853%	−17.206	0.60	0.88
GO:0042623	ATPase activity, coupled	2,501%	−32.190	0.67	0.00
GO:0016462	pyrophosphatase activity	7,130%	−24.489	0.66	0.92
GO:0016818	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	7,190%	−23.585	0.66	0.92
GO:0036402	proteasome-activating ATPase activity	0.005%	−22.947	0.77	0.43
GO:0017111	nucleoside-triphosphatase activity	6774%	−26.352	0.66	0.85
GO:0016887	ATPase activity	4,560%	−18.342	0.67	0.74
GO:0051082	unfolded protein binding	0.486%	−42.881	0.89	0.00
GO:0004457	lactate dehydrogenase activity	0.034%	−19.668	0.92	0.02
GO:0004459	L-lactate dehydrogenase activity	0.018%	−19.668	0.92	0.46
GO:0003755	peptidyl-prolyl cis-trans isomerase activity	0.359%	−15.208	0.91	0.03
GO:0016859	cis-trans isomerase activity	0.364%	−15.208	0.91	0.63
GO:0008144	drug binding	0.178%	−15.902	0.90	0.04
GO:0003743	translation initiation factor activity	0.411%	−18.402	0.85	0.05
GO:1901363	heterocyclic compound binding	41,115%	−170.093	0.81	0.08
GO:0097367	carbohydrate derivative binding	17,252%	−31.349	0.84	0.20
GO:0043168	anion binding	20,942%	−31.876	0.81	0.21
GO:0003676	nucleic acid binding	21,226%	−128.362	0.77	0.21
GO:0036094	small molecule binding	21,337%	−33.401	0.83	0.21
GO:0043167	ion binding	33,492%	−40.525	0.82	0.26
GO:0003723	RNA binding	5,283%	−177.913	0.81	0.29
GO:0097159	organic cyclic compound binding	41,137%	−169.872	0.81	0.29
GO:0032555	purine ribonucleotide binding	16,057%	−34.104	0.64	0.31
GO:0017076	purine nucleotide binding	16,107%	−33.609	0.66	0.70
GO:0030554	adenyl nucleotide binding	14,356%	−15.611	0.65	0.81
GO:0035639	purine ribonucleoside triphosphate binding	15,815%	−32.542	0.65	0.70
GO:0005524	ATP binding	14,125%	−14.318	0.64	0.81
GO:0000166	nucleotide binding	20,185%	−32.657	0.65	0.75
GO:0032559	adenyl ribonucleotide binding	14,309%	−15.755	0.65	0.85
GO:0032553	ribonucleotide binding	16,747%	−33.282	0.65	0.80

(Continued)

TABLE 2 | Continued

	Biological process				
term ID	description	frequency	log$_{10}$ p-value	uniqueness	dispensability
GO:1901265	nucleoside phosphate binding	20,185%	−32.657	0.77	0.33
GO:0016817	hydrolase activity, acting on acid anhydrides	7,223%	−22.918	0.80	0.37

FIGURE 6 | Overlap of DEGs in schistosomula knocked-down for SmJNK and Smp38 MAPKs. Venn diagram representing the number of genes with a decrease **(A)** or increase **(B)** in gene expression after exposure to SmJNK or Smp38 dsRNAs. In red are represented DEGs regulated by Smp38, in blue SmJNK, and in pink are the overlapping DEGs in both datasets.

FIGURE 7 | *C. elegans* RNAi phenotypes enriched in orthologues from down-regulated genes in SmJNK and Smp38 knockdown schistosomula. Worm Enrichr analysis was performed to identify the most significantly enriched RNAi phenotypes listed based on the P-value. The length of the bar represents the significance of that specific term. In addition, the brighter the color, the more significant that term is.

plasminogen activator inhibitor 1 (PAI-1) (Nutter et al., 2015). JNK modulates positively the PAI-1 expression in the neurodegenerative amyloid pathology observed in Alzheimer's disease (Gerenu et al., 2017), whereas JNK inhibition significantly attenuates the induction of PAI-1 (Eurlings et al., 2017). PAI-1 was not detected among DEGs in SmJNK and Smp38 knockdown schistosomula. However, the expression of plasminogen activator inhibitor 1 RNA-binding protein (SERBP1, Smp_009310), which plays a role in regulating the PAI-1 mRNA stability,

was down-regulated (**Figure 4B**). *Schistosoma* inhibits host coagulation during infection through stimulation of fibrinolytic pathways (Mebius et al., 2013). The interaction between parasites and the hosts' fibrinolytic system results in the regulation of several functions related to parasite survival mechanisms. For example, the degradation of immunoglobulins and complement components, the activation of matrix metalloproteinases, the stimulation of adhesion, and the degradation of proteins for nutrition (González-Miguel et al., 2016). The detected reduction

in the SERBP1 transcript levels, could influence PAI-1 stability, therefore, may interfere in parasite feeding and/or reduce mobility in the host, thus, be related to the parasite survival impairment (Andrade et al., 2014).

SmJNK knockdown also down-regulated genes involved in RNA splicing. Alternative splicing events alter the protein repertoire of the cells being regulated by the expression patterns of the splicing factors (Grosso et al., 2008). The splicing control is as complex and relevant as the transcriptional control (Kornblihtt et al., 2004). It has already been shown that JNK signaling pathway controls splicing events in human T cells (Martinez et al., 2015). This pathway also seems to act in the regulation of alternative splicing in response to extracellular stimuli, promoting changes in splicing patterns (Pelisch et al., 2005). Moreover, *Schistosoma* presents trans-splicing, which is a peculiar mechanism that shares the proteins from the spliceosome, could affect up to 58% of transcripts in the cercarial stage (Boroni et al., 2018). It has been suggested as a target for parasite impairment and would be affected by the alterations detected in categories related to RNA splicing (Mourão et al., 2013).

The ability to deal with adverse environmental conditions is a fundamental need for all organisms, especially parasites. It is, therefore, not surprising that stress-activated protein kinase pathways are among the oldest and most conserved metazoan signaling modules (Twumasi-Boateng et al., 2012). The JNK and p38 MAPK pathways share several upstream regulators and, consequently, multiple stimuli simultaneously activate both pathways (Wagner and Nebreda, 2009). Several studies demonstrate that the extracellular signal-regulated kinase (ERK), JNK, and p38 MAPK signaling pathways act cooperatively, amplifying and integrating signals from various stimuli; promoting appropriate physiological responses including cell proliferation, differentiation, development, inflammatory responses, and apoptosis in mammalian cells (Zhang et al., 2002).

Orthologs of SmJNK and Smp38 in *C. elegans* regulate phenotypes that result in sterile nematodes. Hence, oocyte morphology alterations, decreased number or lack of this structure, and gonad alterations are redundant terms among the enriched RNAi phenotypes in the orthologs. The JNK (named kgb-1) deletion in *C. elegans* has a sterile phenotype, characterized by the massive presence of immature oocytes (Smith et al., 2002). The *C. elegans* kgb-1 deleted mutant presents low reproduction rate and shortened lifespan, concomitantly with reduced expression of genes for protein biosynthesis, chaperones, and enzymes involved in ubiquitination/proteasomal degradation (Gerke et al., 2014). All those findings corroborate the phenotypes and transcription profile associated with SmJNK and Smp38 knockdown.

Here we have shown that SmJNK activity is related to the reduction of egg production *in vitro*. This is in accordance with previously observed phenotype alterations *in vivo,* that demonstrated that SmJNK and Smp38 inhibitors could lead to sterile females, thereby reducing schistosomiasis pathology in the host. Our results point to other key regulatory proteins that are not well conserved between host and parasite, encouraging the development of small-molecule inhibitors.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in PRJNA354932 and PRJNA492452.

ETHICS STATEMENT

This work was approved by the Oswaldo Cruz Foundation's Ethics Committee for Animal Use (CEUA) under number LW12/16 , according to the Brazilian national guidelines set out in Law 11794/08.

AUTHOR CONTRIBUTIONS

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.01036/full#supplementary-material

REFERENCES

Andrade, L. F., De, Mourão, M. D. M., Geraldo, J. A., Coelho, F. S., Silva, L. L., Neves, R. H., et al. (2014). Regulation of schistosoma mansoni development and reproduction by the mitogen-activated protein kinase signaling pathway. *PLoS Negl. Trop. Dis.* 8, e2949. doi: 10.1371/journal.pntd.0002949

Andrade, L. F., Nahum, L. A., Avelar, L. G., Silva, L. L., Zerlotini, A., Ruiz, J. C., et al. (2011). Eukaryotic protein kinases (ePKs) of the helminth parasite schistosoma mansoni. *BMC Genomics* 12, 215. doi: 10.1186/1471-2164-12-215

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556

Avelar, L., das, G. A., Gava, S. G., Neves, R. H., Silva, M. C. S., Araújo, N., et al., et al. (2019). Smp38 MAP Kinase Regulation in Schistosoma mansoni: roles in survival, oviposition, and protection against oxidative stress. *Front. Immunol.* 10, 21. doi: 10.3389/fimmu.2019.00021

Bagowski, C. P., Xiong, W., and Ferrell, J. E. (2001). c-Jun N-terminal kinase activation in Xenopus laevis eggs and embryos. A possible non-genomic role for the JNK signaling pathway. *J. Biol. Chem.* 276, 1459–1465. doi: 10.1074/jbc.M008050200

Berriman, M., Haas, B. J., LoVerde, P. T., Wilson, R. A., Dillon, G. P., Cerqueira, G. C., et al. (2009). The genome of the blood fluke Schistosoma mansoni. *Nature* 460, 352–358. doi: 10.1038/nature08160

Bévort, M., and Leffers, H. (2000). Down regulation of ribosomal protein mRNAs during neuronal differentiation of human NTERA2 cells. *Differentiation* 66, 81–92. doi: 10.1046/j.1432-0436.2000.660203.x

Boroni, M., Sammeth, M., Gava, S. G., Jorge, N. A. N., Macedo, A. M., Machado, C. R., et al. (2018). Landscape of the spliced leader trans-splicing mechanism in Schistosoma mansoni. *Sci. Rep.* 8, 3877. doi: 10.1038/s41598-018-22093-3

Bortoluzzi, S., D'Alessi, F., Romualdi, C., and Danieli, G. A. (2001). Differential expression of genes coding for ribosomal proteins in different human tissues. *Bioinformatics* 17, 1152–1157. doi: 10.1093/bioinformatics/17.12.1152

Burke, M. L., Jones, M. K., Gobert, G. N., Li, Y. S., Ellis, M. K., and McManus, D. P. (2009). Immunopathogenesis of human schistosomiasis. *Parasite Immunol.* 31, 163–176. doi: 10.1111/j.1365-3024.2009.01098.x

Cai, P., McManus, D. P., and You, H. (2017). Signalling pathways in schistosomes: novel targets for control interventions against schistosomiasis. *Emerg. Top. Life Sci.* 1, 633–639. doi: 10.1042/ETLS20170093

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G., et al. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* 14, 128. doi: 10.1186/1471-2105-14-128

Cohen, P. (2002). Protein kinases–the major drug targets of the twenty-first century? *Nat. Rev. Drug Discov.* 1, 309–315. doi: 10.1038/nrd773

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635

Espírito Santo, M. C. C., do, Azeredo, L. M., Teles, H. M. S., Gryschek, R. C. B., Ferreira, C. S., and Amato Neto, V. (2008). Abdominal ultrasound in the evaluation of fibrosis and portal hypertension in an area of schistosomiasis low endemicity. *Rev. Inst. Med. Trop. Sao Paulo* 50, 117–119. doi: 10.1590/S0036-46652008000200010

Eurlings, I. M. J. J., Reynaert, N. L., Van De Wetering, C., Aesif, S. W., Mercken, E. M., de Cabo, R., et al. (2017). Involvement of c-Jun N-Terminal Kinase in TNF-α-driven remodeling. *Am. J. Respir. Cell Mol. Biol.* 56, 393–401. doi: 10.1165/rcmb.2015-0195OC

Fielding, A. B., and Dedhar, S. (2009). The mitotic functions of integrin-linked kinase. *Cancer Metastasis Rev.* 28, 99–111. doi: 10.1007/s10555-008-9177-0

French, M. D., Evans, D., Fleming, F. M., Secor, W. E., Biritwum, N.-K., Brooker, S. J., et al. (2018). Schistosomiasis in Africa: improving strategies for long-term and sustainable morbidity control. *PLoS Negl. Trop. Dis.* 12, e0006484. doi: 10.1371/journal.pntd.0006484

Gelmedin, V., Morel, M., Hahnel, S., Cailliau, K., Dissous, C., and Grevelding, C. G. (2017). Evidence for integrin – venus kinase receptor 1 alliance in the ovary of schistosoma mansoni females controlling cell survival. *PLoS Pathog.* 13, e1006147. doi: 10.1371/journal.ppat.1006147

Gerenu, G., Martisova, E., Ferrero, H., Carracedo, M., Rantamäki, T., Ramirez, M. J., et al. (2017). Modulation of BDNF cleavage by plasminogen-activator inhibitor-1 contributes to Alzheimer's neuropathology and cognitive deficits.

Biochim. Biophys. Acta Mol. Basis Dis. 1863, 991–1001. doi: 10.1016/j.bbadis.2017.01.023

Gerke, P., Keshet, A., Mertenskötter, A., and Paul, R. J. (2014). The JNK-Like MAPK KGB-1 of Caenorhabditis elegans promotes reproduction, lifespan, and gene expressions for protein biosynthesis and germline homeostasis but interferes with hyperosmotic stress tolerance. *Cell Physiol. Biochem.* 34, 1951–1973. doi: 10.1159/000366392

González-Miguel, J., Siles-Lucas, M., Kartashev, V., Morchón, R., and Simón, F. (2016). Plasmin in parasitic chronic infections: friend or Foe? *Trends Parasitol.* 32, 325–335. doi: 10.1016/j.pt.2015.12.012

Grevelding, C. G., Langner, S., and Dissous, C. (2017). Kinases: Molecular Stage directors for schistosome development and differentiation. *Trends Parasitol.* xx, 1–15. doi: 10.1016/j.pt.2017.12.001

Grosso, A. R., Gomes, A. Q., Barbosa-Morais, N. L., Caldeira, S., Thorne, N. P., Grech, G., et al. (2008). Tissue-specific splicing factor gene expression signatures. *Nucleic Acids Res.* 36, 4823–4832. doi: 10.1093/nar/gkn463

Howe, K. L., Bolt, B. J., Shafie, M., Kersey, P., and Berriman, M. (2017). WormBase ParaSite - a comprehensive resource for helminth genomics. *Mol. Biochem. Parasitol.* 215, 2–10. doi: 10.1016/j.molbiopara.2016.11.005

Huang, X., Tong, J., Wang, Z., Yang, C., Qi, S.-T., Guo, L., et al. (2011). JNK2 Participates in Spindle Assembly during Mouse Oocyte Meiotic Maturation. *Microsc. Microanal.* 17, 197–205. doi: 10.1017/S1431927610094456

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45, D353–D361. doi: 10.1093/nar/gkw1092

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/27.1.29

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457–D462. doi: 10.1093/nar/gkv1070

King, C. H., and Dangerfield-Cha, M. (2008). The unacknowledged impact of chronic schistosomiasis. *Chronic Illn.* 4, 65–79. doi: 10.1177/1742395307084407

Kornblihtt, A. R., de la Mata, M., Fededa, J. P., Munoz, M. J., and Nogues, G. (2004). Multiple links between transcription and splicing. *RNA* 10, 1489–1498. doi: 10.1261/rna.7100104

Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 44, W90–W97. doi: 10.1093/nar/gkw377

Lee, C., Kim, J., Shin, S. G., and Hwang, S. (2006). Absolute and relative QPCR quantification of plasmid copy number in Escherichia coli. *J. Biotechnol.* 123, 273–280. doi: 10.1016/j.jbiotec.2005.11.014

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 1–21. doi: 10.1186/s13059-014-0550-8

Lu, Z., Sessler, F., Holroyd, N., Hahnel, S., Quack, T., Berriman, M., et al. (2016). Schistosome sex matters: a deep view into gonad-specific and pairing-dependent transcriptomes reveals a complex gender interplay. *Sci. Rep.* 6, 31150. doi: 10.1038/srep31150

Lu, Z., Sessler, F., Holroyd, N., Hahnel, S., Quack, T., Berriman, M., et al. (2017). Data Descriptor: a gene expression atlas of adult Schistosoma mansoni and their gonads. *Sci. Data* 4, 1–9. doi: 10.1038/sdata.2017.118

Mager, W. H. (1988). Control of ribosomal protein gene expression. *Biochim. Biophys. Acta* 949, 1–15. doi: 10.1016/0167-4781(88)90048-6

Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002). The protein kinase complement of the human genome. *Science* 298, 1912–34. doi: 10.1126/science.1075762

Marcellino, C., Gut, J., Lim, K. C., Singh, R., McKerrow, J., and Sakanari, J. (2012). WormAssay: a novel computer application for whole-plate motion-based screening of macroscopic parasites. *PLoS Negl. Trop. Dis.* 6, e1494. doi: 10.1371/journal.pntd.0001494

Martinez, N. M., Agosto, L., Qiu, J., Mallory, M. J., Gazzara, M. R., Barash, Y., et al. (2015). Widespread JNK-dependent alternative splicing induces a positive feedback loop through CELF2-mediated regulation of MKK7 during T-cell activation. *Genes Dev.* 29, 2054–2066. doi: 10.1101/gad.267245.115

SmJNK Transcriptional Regulation in *Schistosoma*

Mebius, M. M., van Genderen, P. J. J., Urbanus, R. T., Tielens, A. G. M., de Groot, P. G., and van Hellemond, J. J. (2013). Interference with the host haemostatic system by schistosomes. *PLoS Pathog.* 9, 1–8. doi: 10.1371/journal.ppat.1003781

Milligan, J. N., and Jolly, E. R. (2011). Cercarial transformation and in vitro cultivation of Schistosoma mansoni schistosomules. *J. Vis. Exp.* 54, 4–7. doi: 10.3791/3191

Mourão, M., de, M., Bitar, M., Lobo, F. P., Peconick, A. P., Grynberg, P., et al., et al. (2013). A directed approach for the identification of transcripts harbouring the spliced leader sequence and the effect of trans-splicing knockdown in Schistosoma mansoni. *Mem. Inst. Oswaldo Cruz* 108, 707–717. doi: 10.1590/0074-0276108062013006

Mourão, M. M., de Moraes Mourão, M., Dinguirard, N., Franco, G. R., and Yoshino, T. P. (2009). Phenotypic screen of early-developing larvae of the blood fluke, Schistosoma mansoni, using RNA interference. *PLoS Negl. Trop. Dis.* 3, e502. doi: 10.1371/journal.pntd.0000502

Nagiah, S., Phulukdaree, A., and Chuturgoon, A. A. (2016). Lon protease and eiF2α are involved in acute, but not prolonged, antiretroviral induced stress response in HepG2 cells. *Chem. Biol. Interact.* 252, 82–86. doi: 10.1016/j.cbi.2016.03.021

Nascimento, G. L., Pegado, H. M., Domingues, A. L. C., Ximenes, R. A., de, A., Itria, A., et al., et al. (2019). The cost of a disease targeted for elimination in Brazil: the case of schistosomiasis mansoni. *Mem. Inst. Oswaldo Cruz* 114, 1–9. doi: 10.1590/0074-02760180347

Nutter, F. H., Haylor, J. L., and Khwaja, A. (2015). Inhibiting ERK activation with CI-1040 leads to compensatory upregulation of alternate MAPKs and plasminogen activator inhibitor- 1 following subtotal nephrectomy with no impact on kidney fibrosis. *PLoS One* 10, 1–18. doi: 10.1371/journal.pone.0137321

Pelisch, F., Blaustein, M., Kornblihtt, A. R., and Srebrow, A. (2005). Cross-talk between signaling pathways regulates alternative splicing: A novel role for JNK. *J. Biol. Chem.* 280, 25461–25469. doi: 10.1074/jbc.M412007200

Pellegrino, J., and Siqueira, A. F. (1956). A perfusion technic for recovery of Schistosoma mansoni from experimentally infected guinea pigs. *Rev. Bras. Malariol. Doencas Trop.* 8, 589–597.

R Development Core Team (2011). R: A language and environment for statistical computing.

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., et al. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* 47, W191–W198. doi: 10.1093/nar/gkz369

Reimand, J., Kull, M., Peterson, H., Hansen, J., and Vilo, J. (2007). g:Profiler–a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res.* 35, W193–W200. doi: 10.1093/nar/gkm226

Smith, P., Leung-Chiu, W. M., Montgomery, R., Orsborn, A., Kuznicki, K., Gressman-Coberly, E., et al. (2002). The GLH proteins, Caenorhabditis elegans P granule components, associate with CSN-5 and KGB-1, proteins necessary for fertility, and with ZYX-1, a predicted cytoskeletal protein. *Dev. Biol.* 251, 333–347. doi: 10.1006/dbio.2002.0832

Stroehlein, A. J., Young, N. D., Jex, A. R., Sternberg, P. W., Tan, P., Boag, P. R., et al. (2015). Defining the Schistosoma haematobium kinome enables the prediction of essential kinases as anti-schistosome drug targets. *Sci. Rep.* 5, 17759. doi: 10.1038/srep17759

Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6, e21800. doi: 10.1371/journal.pone.0021800

The Gene Ontology Consortium (2017). Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.* 45, D331–D338. doi: 10.1093/nar/gkw1108

Twumasi-Boateng, K., Wang, T. W., Tsai, L., Lee, K. H., Salehpour, A., Bhat, S., et al. (2012). An age-dependent reversal in the protective capacities of JNK signaling shortens Caenorhabditis elegans lifespan. *Aging Cell* 11, 659–667. doi: 10.1111/j.1474-9726.2012.00829.x

Wagner, E. F., and Nebreda, A. R. (2009). Signal integration by JNK and p38 MAPK pathways in cancer development. *Nat. Rev. Cancer* 9, 537–549. doi: 10.1038/nrc2694

Wang, J., Lan, P., Gao, H., Zheng, L., Li, W., and Schmidt, W. (2013). Expression changes of ribosomal proteins in phosphate- and iron-deficient Arabidopsis roots predict stress-specific alterations in ribosome composition. *BMC Genomics* 14, 783. doi: 10.1186/1471-2164-14-783

WHO (2018). Weekly epidemiological record:Schistosomiasis and soil-transmitted helminthiases: numbers of people treated in 2017. *Wkly. Epidemiol. Rec.* III, 445–452. doi: 10.1016/j.actatropica.2012.04.013

Zhang, W., Liu, H. T., and Tu LIU, H. (2002). MAPK signal pathways in the regulation of cell proliferation in mammalian cells. *Cell Res.* 12, 9–18. doi: 10.1038/sj.cr.7290105

Complex I and II Subunit Gene Duplications Provide Increased Fitness to Worms

Lucía Otero[1†], Cecilia Martínez-Rosales[1†], Exequiel Barrera[2], Sergio Pantano[2] and Gustavo Salinas[1*]

[1] Laboratorio de Biología de Gusanos, Unidad Mixta Departamento de Biociencias, Facultad de Química, Universidad de la República–Institut Pasteur de Montevideo, Montevideo, Uruguay, [2] Laboratorio de Simulaciones Biomoleculares, Institut Pasteur de Montevideo, Montevideo, Uruguay

Helminths use an alternative mitochondrial electron transport chain (ETC) under hypoxic conditions, such as those found in the gastrointestinal tract. In this alternative ETC, fumarate is the final electron acceptor and rhodoquinone (RQ) serves as an electron carrier. RQ receives electrons from reduced nicotinamide adenine dinucleotide through complex I and donates electrons to fumarate through complex II. In this latter reaction, complex II functions in the opposite direction to the conventional ETC (i.e., as fumarate reductase instead of succinate dehydrogenase). Studies in *Ascaris suum* indicate that this is possible due to changes in complex II, involving alternative succinate dehydrogenase (SDH) subunits SDHA and SDHD, derived from duplicated genes. We analyzed helminth genomes and found that distinct lineages have different gene duplications of complex II subunits (SDHA, SDHB, SDHC, and SDHD). Similarly, we found lineage-specific duplications in genes encoding complex I subunits that interact with quinones (NDUF2 and NDUF7). The phylogenetic analysis of ETC subunits revealed a complex history with independent evolutionary events involving gene duplications and losses. Our results indicated that there is not a common evolutionary event related to ETC subunit genes linked to RQ. The free-living nematode *Caenorhabditis elegans* uses RQ and has two genes encoding SDHA (*sdha-1* and *sdha-2*) and two genes encoding NDUF2 (*nduf2-1* and *nduf2-2*). *sdha-1* and *nduf2-1* are essential genes and have a similar expression pattern during *C. elegans* lifecycle. Using knockout strains, we found that *sdha-2* and *nduf2-2* are not essential, even in hypoxia. Yet, *sdha-2* and *nduf2-2* expression is increased in the early embryo and in dauer larvae, stages where there is low oxygen tension. Strikingly, *sdha-1* and *sdha-2* as well as *nduf2-1* and *nduf2-2* showed inverted expression profiles during the *C. elegans* life cycle. Finally, we found that *sdha-2* and *nduf2-2* knockout mutant strain progeny is affected. Our results indicate that different complex I and II subunit gene duplications provide increased fitness to worms.

Keywords: rhodoquinone, *C. elegans*, nematode, platyhelminth, electron transport chain, hypoxia, helminth, gas-1
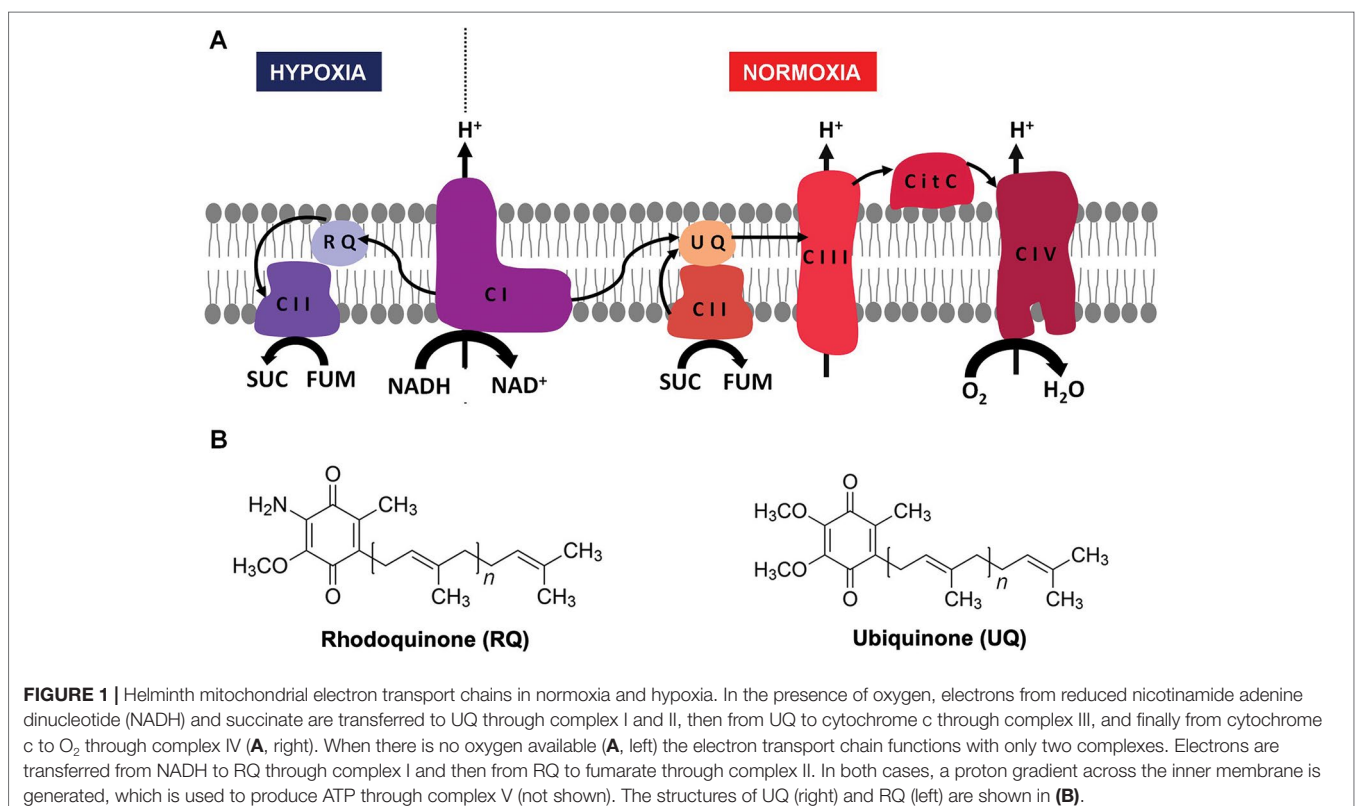

Abbreviations: FRD, fumarate reductase; RQ, rhodoquinone; SDH, succinate dehydrogenase; UQ, ubiquinone.

INTRODUCTION

Energy is essential to all forms of life. Helminths live part of their life cycle under hypoxic conditions and have an energy metabolism different from their hosts. Under hypoxia, helminths use fumarate instead of oxygen as a final electron acceptor of the electron transport chain (ETC) (**Figure 1A**) (Hellemond et al., 2003; Harada et al., 2013). This is achieved by means of an alternative ETC, in which rhodoquinone (RQ) serves as an electron carrier (Van Hellemond et al., 1995; Hellemond et al., 2003; Harada et al., 2013). RQ is a redox-active quinone structurally similar to the conventional ubiquinone (UQ), differing only in the 6-methoxy substituent of UQ that is changed to an amino substituent in RQ (**Figure 1B**) (Moore and Folkers, 1965). This change confers RQ a lower redox potential than UQ (-63 and 43 mV, respectively) (Erabi et al., 1975). This allows RQ to receive electrons from reduced nicotinamide adenine dinucleotide (NADH) through complex I and to reduce fumarate to succinate (redox potential 30 mV) through complex II. Thus, in this alternative ETC, complex II functions as a fumarate reductase (FRD), in the opposite direction to the conventional ETC in which complex II functions as succinate dehydrogenase (SDH) (Van Hellemond et al., 1995)(Tielens and Van Hellemond, 1998). SDH oxidizes succinate to fumarate and reduces UQ. Importantly, the alternative ETC also allows proton pumping through complex I. The resulting proton gradient is coupled to ATP synthesis by mitochondrial ATP synthase. The fumarate and NADH used in this ETC are generated in the malate dismutation pathway (Tielens and Van Hellemond, 1998). This pathway yields

between 5 and 6 mol of ATP per mol of glucose. A similar ETC that uses RQ is also present in some prokaryotes, protists, and other animals that alternate cycles of normoxia and hypoxia, such as bivalves and freshwater snails (Van Hellemond et al., 1995).

RQ has been found in all helminths where its presence has been examined, being more abundant in stages that dwell in hypoxic environments (Sato and Ozawa, 1969; Allen, 1973; Van Hellemond et al., 1995). Importantly, RQ has also been associated with hypoxia in the free-living nematode *Caenorhabditis elegans* (Takamiya et al., 1999), an amenable model for genetic studies. It is thought that the use RQ as an electron carrier requires adjustments in ETC complex II subunits. Several studies in *Ascaris suum*, a nematode in which biochemical studies are feasible, have shown that the use of RQ is possible due to changes in complex II subunits (Takamiya et al., 1993; Kuramochi et al., 1994; Iwata et al., 2008; Amino et al., 2003). Mitochondria from aerobic developmental stages of *A. suum*, larval stages L2 and L3, contain mainly UQ. In contrast, mitochondria from adult worms, which live in an oxygen-deprived environment, contain only RQ (Takamiya et al., 1993). The transition from normoxia to hypoxia is accompanied by an exchange of the SDHA and SDHD subunits of complex II, leading to the change in enzymatic activities of complex II, from SDH-UQ reductase to FRD-rhodoquinol oxidase (Takamiya et al., 1993; Saruta et al., 1995; Amino et al., 2003; Iwata et al., 2008). In the nematode *Haemonchus contortus*, two genes encoding distinct SDHB subunits were identified and showed different expression patterns during development (Roos and Tielens, 1994). This phenomenon has also been linked to a metabolic transition due to a change in oxygen tension. In addition, FRD and SDH activities

FIGURE 1 | Helminth mitochondrial electron transport chains in normoxia and hypoxia. In the presence of oxygen, electrons from reduced nicotinamide adenine dinucleotide (NADH) and succinate are transferred to UQ through complex I and II, then from UQ to cytochrome c through complex III, and finally from cytochrome c to O_2 through complex IV (**A**, right). When there is no oxygen available (**A**, left) the electron transport chain functions with only two complexes. Electrons are transferred from NADH to RQ through complex I and then from RQ to fumarate through complex II. In both cases, a proton gradient across the inner membrane is generated, which is used to produce ATP through complex V (not shown). The structures of UQ (right) and RQ (left) are shown in **(B)**.

have been detected in other helminths (Fioravanti et al., 1998; Matsumoto et al., 2008), but the complex II composition has not been examined. In contrast to complex II, no transitions associated with oxygen tension have been reported in helminth complex I (NADH dehydrogenase). The genes encoding ETC complexes I and II have not been studied in detail after helminth genomes have been made publicly available.

We examined nematode and platyhelminth genomes and found diverse gene duplications in complex II subunits. Furthermore, gene duplications in complex I subunits that interact with quinones were identified in helminth genomes. Using *C. elegans*, we studied the role of complex I and II subunits duplications.

RESULTS

Different Helminths Have Dissimilar Complex I and II Subunit Duplications
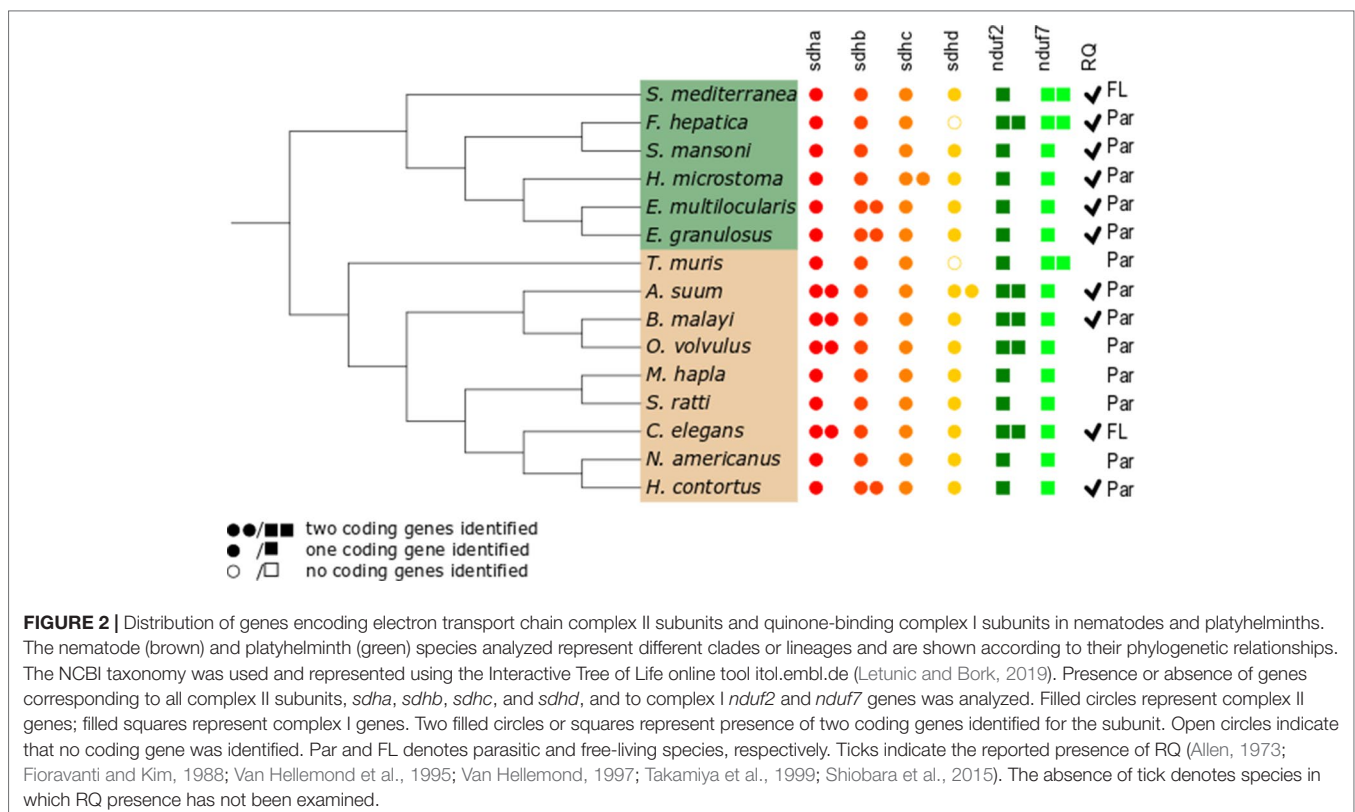
Since gene duplications of complex II subunits have been described in some helminths, we examined the genes encoding complex II in representative and well-annotated parasitic and free-living nematode and platyhelminth genomes. Interestingly, we observed diversity in complex II genes. The absence/presence of complex II genes is presented in **Figure 2** (alignments for all subunits studied are available in **Supplementary Figure 1**).
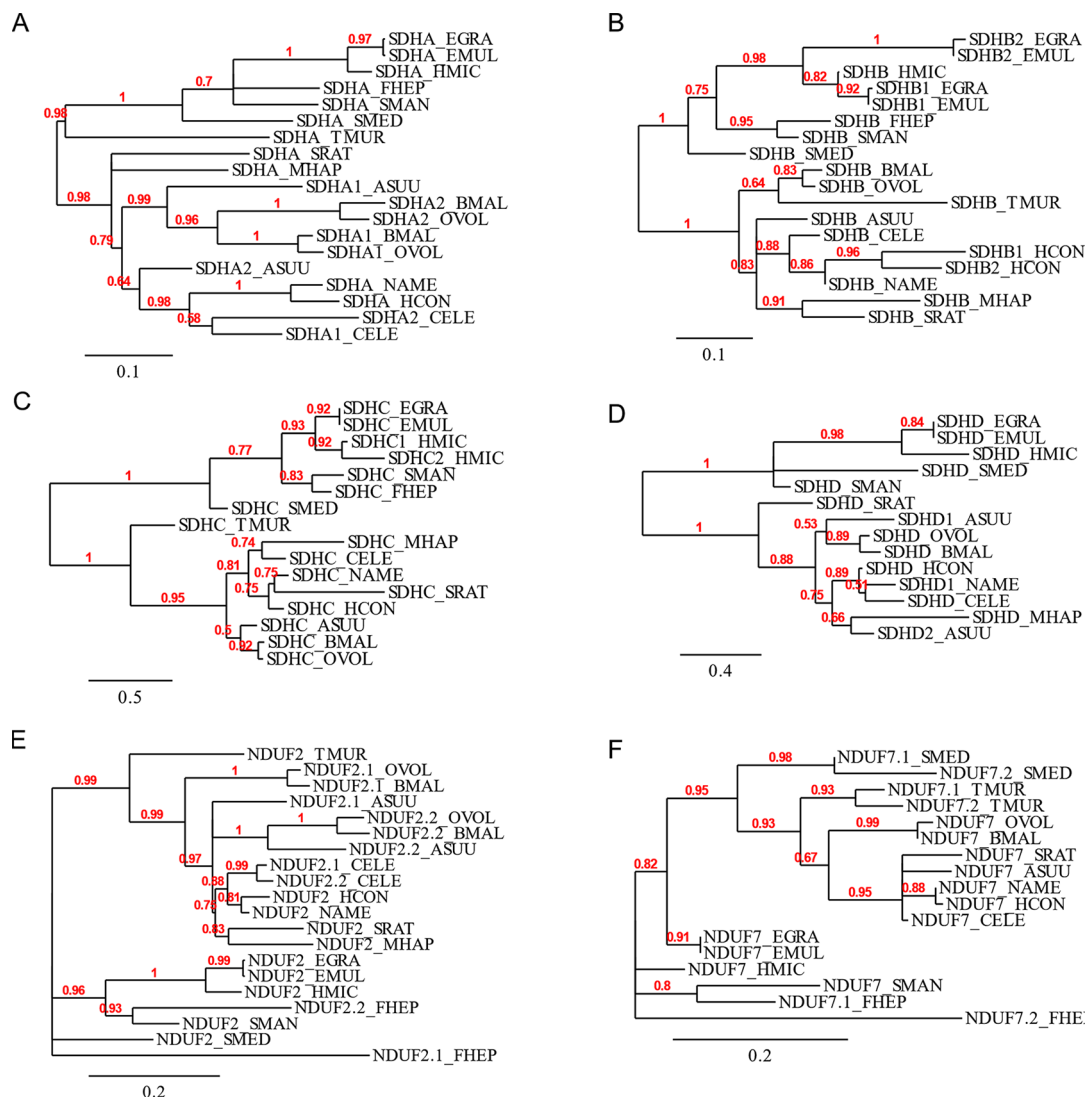
The flavoprotein subunit (SDHA) gene was found to be duplicated in some, but not all, nematodes, whereas this gene duplication is not observed in any platyhelminth lineage

(**Figure 2**). The phylogenetic analysis does not indicate a clear evolutionary event associated with *sdha* history (**Figure 3A**) and suggests that independent gene duplications have occurred in filarial and *C. elegans* lineages. Indeed, filarial SDHA-1 and SDHA-2 form different nodes, and they cluster together in a putative ancestral filarial SDHA node. Likewise, *C. elegans* SDHA-1 and SDHA-2 form a node, indicating a recent gene duplication event in this lineage. However, in the case of *A. suum*, the two SDHAs do not group together, suggesting that the two SDHA subunits diverged extensively after duplication. Alternatively, it may represent an ancient gene duplication that was subsequently lost in other lineages. This latter scenario is less parsimonious since it would imply several gene losses and subsequent duplications.

Two genes encoding alternative iron–sulfur subunits (SDHB) have been previously described in *H. contortus* (Roos and Tielens, 1994). This gene duplication is absent in all other nematode genomes analyzed. Interestingly, we found that *Echinococcus* spp. also exhibit an *sdhb* duplication in the platyhelminth lineage. The phylogenetic analysis of SDHB suggests independent gene duplications in these lineages.

Regarding the heme-containing membrane subunits (SDHC and SDHD), the results were unexpected. *sdhc* was found to be duplicated only in *Hymenolepis microstoma*, whereas *sdhd* was found to be duplicated only in *A. suum*. Strikingly, *Fasciola hepatica* and *Trichuris muris* lost *sdhd*, suggesting independent gene losses in these lineages. It is important to note that transcripts corresponding to *sdhd* were absent in these species'

FIGURE 2 | Distribution of genes encoding electron transport chain complex II subunits and quinone-binding complex I subunits in nematodes and platyhelminths. The nematode (brown) and platyhelminth (green) species analyzed represent different clades or lineages and are shown according to their phylogenetic relationships. The NCBI taxonomy was used and represented using the Interactive Tree of Life online tool itol.embl.de (Letunic and Bork, 2019). Presence or absence of genes corresponding to all complex II subunits, *sdha*, *sdhb*, *sdhc*, and *sdhd*, and to complex I *nduf2* and *nduf7* genes was analyzed. Filled circles represent complex II genes; filled squares represent complex I genes. Two filled circles or squares represent presence of two coding genes identified for the subunit. Open circles indicate that no coding gene was identified. Par and FL denotes parasitic and free-living species, respectively. Ticks indicate the reported presence of RQ (Allen, 1973; Fioravanti and Kim, 1988; Van Hellemond et al., 1995; Van Hellemond, 1997; Takamiya et al., 1999; Shiobara et al., 2015). The absence of tick denotes species in which RQ presence has not been examined.

FIGURE 3 | Phylogenetic trees of complex II and nuclear-encoded quinone-binding complex I subunits. Phylograms of **(A)** SDHA, **(B)** SDHB, **(C)** SDHC, **(D)** SDHD, **(E)** NDUF2, and **(F)** NDUF7. Multiple sequence alignments were performed with MUSCLE 3.8.31 and curated with Gblocks 0.91b. Phylograms were inferred using PhyML 3.0 from http://www.phylogeny.fr. The values above each branch indicate bootstrap values. Polytomies denote nodes for which the relationships cannot be resolved to dichotomies. Alignments are shown in **Supplementary Figure 1**.

transcriptomes. Since ETC subunit genes are highly expressed, the absence of *sdhd* transcripts favors the hypothesis that this gene is absent in these genomes.

Since adjustments in complex II are thought to be related to the use of alternative quinones (UQ or RQ), we also analyzed complex I subunits NDUF2 and NDUF7, involved in quinone binding. Interestingly, we found that the same nematodes that have two *sdha* genes have also two *nduf2* genes. In platyhelminths, only *F. hepatica* has two *nduf2* genes. Phylogenetic analysis of NDUF2 in nematodes suggests independent events of gene duplication in *C. elegans* and in clade III lineages. In the case of platyhelminth NDUF2, it is likely that a gene duplication event occurred in the *F. hepatica* lineage and rapidly diverged. However, an ancestral gene duplication (e.g., at the root of trematodes) followed by gene loss in independent lineages cannot be ruled out.

We also found that *T. muris*, *Schimdtea mediterranea*, and *F. hepatica* have two *nduf7*. According to the phylogenetic tree, the gene duplications observed in *T. muris* and *S. mediterranea* lineages appear to be recent events. Thus, the most likely explanation is that these gene duplication events occurred independently.

The analysis of the ETC subunit genes revealed a complex history with independent evolutionary events without an evident pattern. In the case of *A. suum*, the different complex II subunits used in the larval and adult stages have been associated with oxygen tension changes found during its lifecycle (normoxia and hypoxia, respectively) and the use of alternative quinones (UQ and RQ, respectively). However, a conclusion supported by our study is the absence of a common evolutionary event related to ETC subunit genes directly linked to RQ, the key metabolite during hypoxia.

The Quinone-Binding Site in Normoxia and Hypoxia Complex II Is Highly Conserved

We performed *in silico* docking analysis in order to explore how the change in subunits of complex II may affect the potential binding modes of different quinones. To this end, three structures were used: the crystal structure of *A. suum* complex II purified from adult worms (PDB: 5C2T) and homology models of complex II obtained for *A. suum* larval stages and *C. elegans*. The latter encodes one gene for each subunit involved in binding quinones (SDHB, SDHC, and SDHD). The search space, indicated by a box with black edges in **Figure 4A**, was defined considering the binding site of RQ_2 (*i.e.* RQ with two isoprene units) in 5C2T and giving sufficient volume to explore alternative binding modes. Multiple sequence alignments were performed on the SDH subunits of *A. suum* (adult and larval stages) and *C. elegans*. Results indicated that, within the docking search space, only eight amino acids are not conserved (six for SDHC and two for SDHD, **Supplementary Figure 2**). Those residues face oppositely to the ligand and are not involved in direct interactions with RQ_2 (**Figure 4B**).

Docking calculations employing the three complex II models showed that the binding modes with higher affinities for both

FIGURE 4 | Complex II quinone-binding site. **(A)** Cartoon representation of the adult *A. suum* complex II crystallized structure bound to RQ_2 (i.e., with two isoprenoid units). Color code: Purple = SDHA; red = SDHB; orange = SDHC; yellow = SDHD. **(B)** Inset showing the docking search space (black lines), RQ_2 (spheres and sticks representation), and nonconserved residues among *A. suum* and *C. elegans* included in the search space (tubes). **(C)** Comparison between the best docking solution for RQ_2 (cyan spheres and sticks) and its binding site obtained from X-ray crystal structure (transparent surface). The best docking solution showed a root mean square deviation (RMSD) of 0.15 nm. **(D)** Superimposition of the best docking solutions for UQ_2 and RQ_2 performed for *C. elegans* and both adult and larva *A. suum* complex II models (transparent surface).

UQ_2 and RQ_2 fit without any steric impediment into the quinone binding pocket (**Figure 4D**). These results suggest that SDHB, SDHC, and SDHD gene duplications would not reflect an adaptation to the binding site of UQ or RQ.

We then analyzed the residues involved in complex II quinone binding in nematodes and platyhelminths. We found that the residues of SDHB and SDHD are conserved in all the analyzed species. SDHC allows some changes in quinone-binding positions (**Figure 5**). Interestingly, in both *H. microstoma* SDHC, the only species that has two *sdhc*, the quinone-binding positions are conserved.

sdha-1 and *sdha-2* as Well as *nduf2-1 and nduf-2.2* Showed an Inverted Expression Profile During *C. elegans* Lifecycle

In the *C. elegans* genome, there are two *sdha* and two *nduf2* genes, as in *A. suum*. *C. elegans* SDHA-1 is essential and has the highest level of homology to *A. suum* SDHA-1 (larval SDHA) and *C. elegans* SDHA-2 to *A. suum* SDHA-2 (adult SDHA). In the case of NDUF2, NDUF2-1 is essential, and both *C. elegans* proteins (NDUF2-1 and NDUF2-2) are more similar to each other than to any NDUF2 proteins of *A. suum*. We analyzed the expression of *sdha* and *nduf2* during *C. elegans* lifecycle. The results are shown in **Figure 6**. In general, *sdha-1* has a higher expression than *sdha-2*, and both genes show an inverted expression profile along the different developmental stages. *sdha-1* expression increases towards the end of the embryonic development and is highest during larval stages L2–L3 and then decreases in the last larval stage (L4) and in the adult worm (**Figure 6A**). In contrast, *sdha-2* expression is highest during the early embryo and then starts to decrease near L1 larval stage. Its expression increases again towards the adult worm (**Figure 6A**). *C. elegans* has an alternative larval stage named dauer, which allows it to endure when environmental conditions are not adequate for normal growth (Cassada and Russell, 1975). After entry in dauer, *sdha-1* expression decreases while *sdha-2* expression increases (**Figure 6A**).
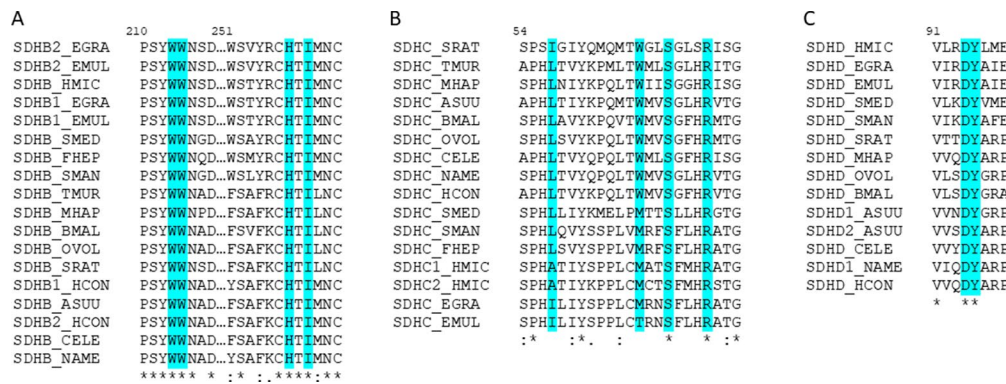
Similar to *sdha-1* and *sdha-2*, *nduf2-1* expression is higher than *nduf2-2* along *C. elegans* lifecycle, except at the beginning of the embryo development. The expression patterns of *nduf2-1* and *nduf2-2* echo that of *sdha-1* and *sdha-2*, respectively, in the embryo, larval and adult stages, as well as in the dauer larvae (**Figure 6B**). Interestingly, during embryonic development, *nduf2-1* and *nduf2-2* expression curves cross each other around the time of egg laying.

These observations suggest that there is a tight regulation of the expression of SDHA and NDUF2 coding genes throughout the lifecycle of *C. elegans*.
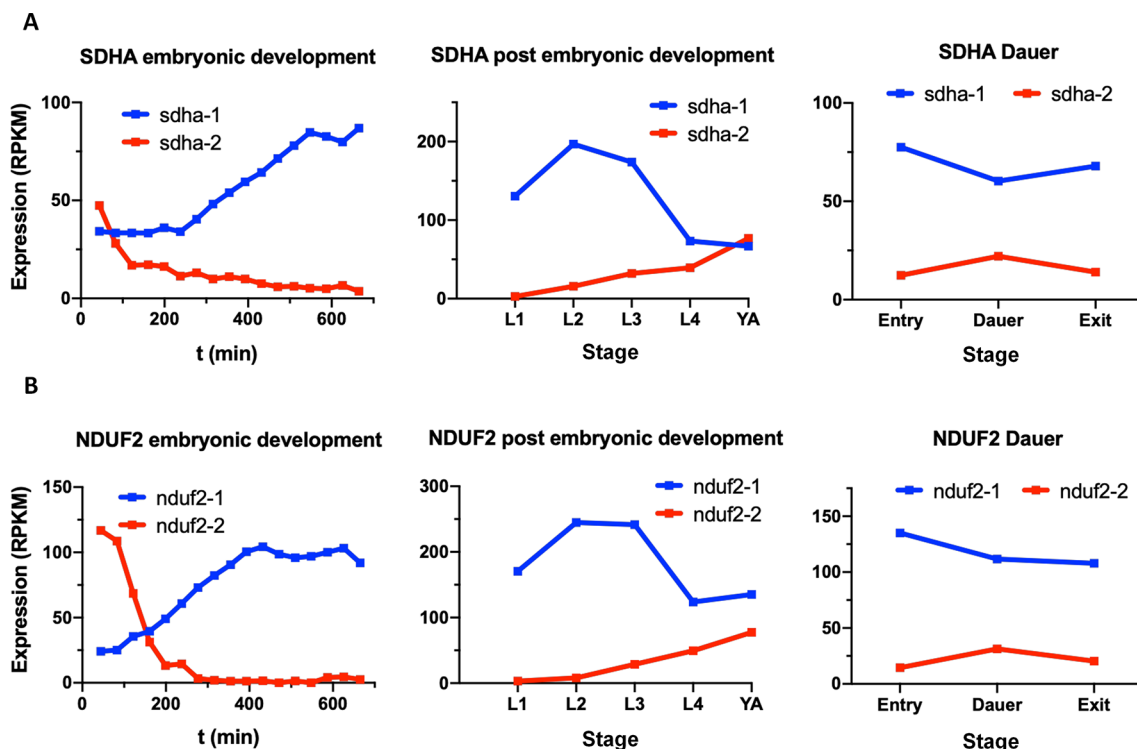
Progeny Is Reduced in *sdha-2 and nduf2-2* Mutant Worms

As already mentioned, the expression of *nduf2-2* and *sdha-2* increases in the adult worm and is highest in the early embryo. Thus, we assessed the effect of the absence of these genes on the reproductive potential of the worm. The analysis of the brood size of KO strains in *nduf2-2* and *sdha-2* showed that
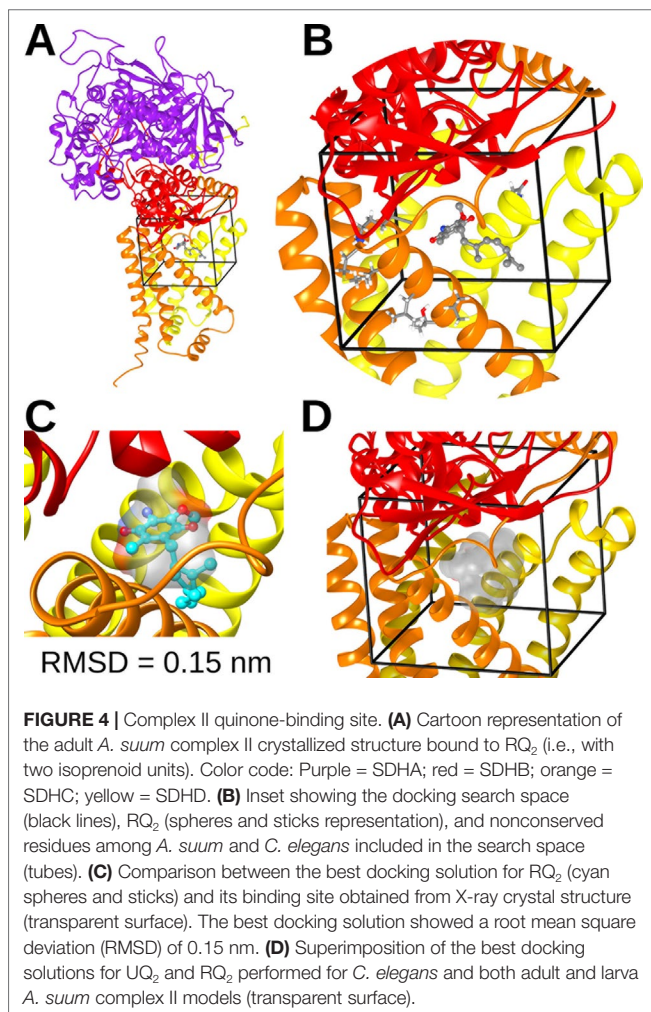
FIGURE 5 | The amino acids interacting with quinones in complex II are highly conserved. Alignment of helminths' complex II sequences containing the identified amino acids involved in quinone-binding in **(A)** SDHB, **(B)** SDHC, and **(C)** SDHD. In turquoise are highlighted the residues identified as interacting with quinones. The numeration corresponds to *C. elegans* sequences.
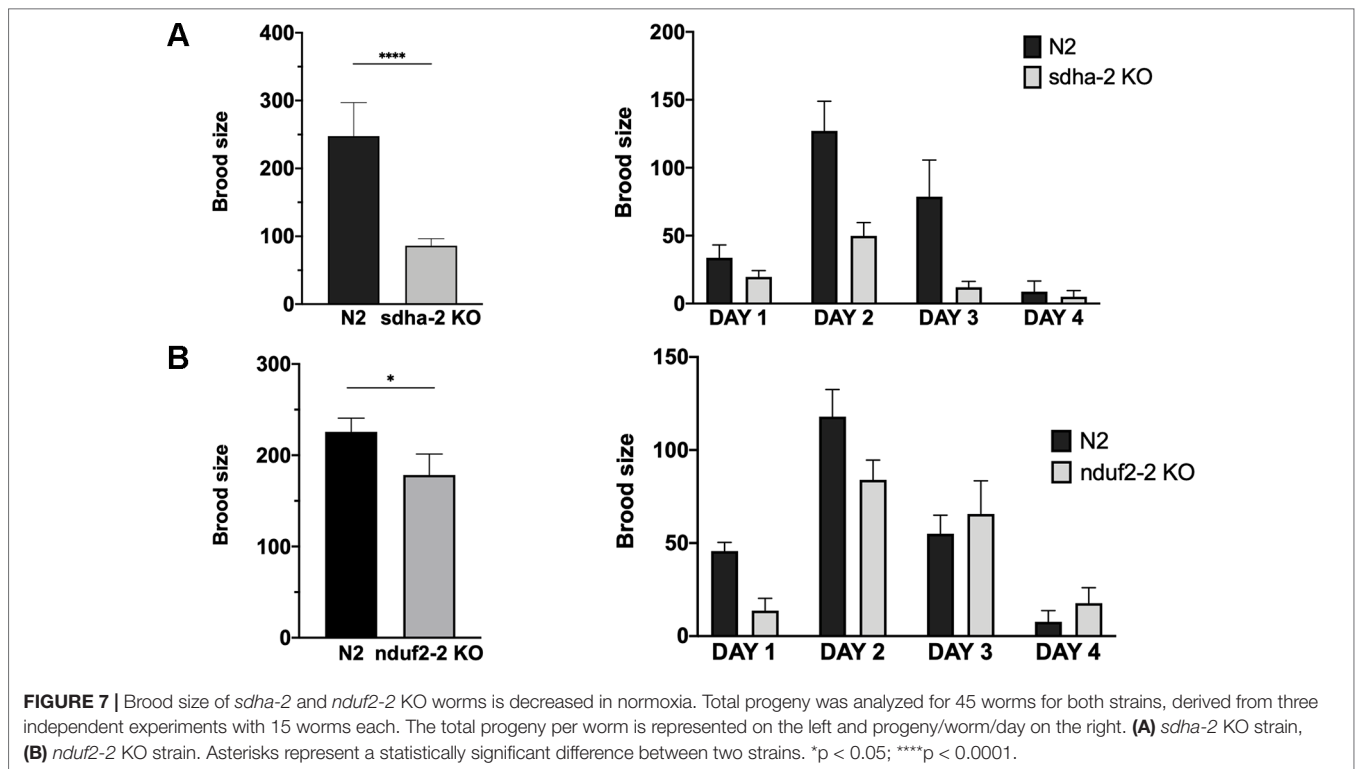

FIGURE 6 | Expression of *sdha* and *nduf2* genes throughout *C. elegans* lifecycle. Expression of *sdha-1* and *sdha-2* **(A)**, and *nduf2-1* and *nduf2-2* **(B)** during *C. elegans* development. Expression during embryonic development is shown on the left graphs. Egg laying occurs around 150 min after fertilization. The four larval stages (L1, L2, L3, and L4) and young adults are shown in the middle graphs. The expression levels for dauer larvae at three timepoints (entry, dauer, and exit) are shown on the right graphs.

the total amount of offspring is significantly reduced in both KO worms compared to the wild-type strain (N2) (**Figure 7**); this reduction was more pronounced in the *sdha-2* KO worms. We also observed a few hours delay in development from L1 to the beginning of egg laying in *nduf2-2* KO worms compared to N2.

Since in *A. suum* a duplication in the SDHA coding gene has been associated with an adaptation to environments with

different oxygen tensions, we examined whether the duplications in *sdha* and *nduf2* in *C. elegans* could also be related to changes in oxygen availability. Similar to N2, *nduf2-2* and *sdha-2* KO strains survived after 24 h exposure to hypoxia (0.4% O_2) or anoxia. We then tested brood size for these strains after a 24-h exposure to hypoxia or anoxia. The phenotype previously observed under normoxic conditions was not exacerbated under hypoxia or anoxia for these mutants (**Supplementary Figure 4**).

FIGURE 7 | Brood size of *sdha-2* and *nduf2-2* KO worms is decreased in normoxia. Total progeny was analyzed for 45 worms for both strains, derived from three independent experiments with 15 worms each. The total progeny per worm is represented on the left and progeny/worm/day on the right. **(A)** *sdha-2* KO strain, **(B)** *nduf2-2* KO strain. Asterisks represent a statistically significant difference between two strains. *$p < 0.05$; ****$p < 0.0001$.

Our results indicate that *sdha-2* and *nduf2-2* are not essential for fecundity, but the absence of these genes reduces progeny production of the worm.

DISCUSSION

RQ has been identified in all nematodes and platyhelminths that have been analyzed for its presence. Owing to its redox potential (intermediate between complex I and II), RQ is thought to be the key metabolite in the alternative ETC used under hypoxic conditions. Until now, it was thought that the use of RQ in helminths required alternative complex II subunits derived from duplicated genes. Our results contradict this view. Indeed, some species do not have any complex II gene duplications (e.g., *Meloidogyne hapla*, *Strongyloides ratti*, *Necator americanus*, *S. mediterranea*, and *Schistosoma mansoni*), while others have even lost *sdhd* (*F. hepatica* and *T. muris*). Furthermore, there is not a unique evolutionary event associated with the use of RQ in complex I and II genes. We examined several life history traits for the different species whose genomes were analyzed. In particular, we examined absence/presence of free-living stage(s), parasitized host(s) (intermediate, definitive, vector), parasitized site/tissue, and reproduction mode (monoecious vs. dioecious). We have not found any particular association of life history traits or style that could explain a trend regarding the presence of alternative ETCs in the different lineages. Thus, RQ has not been a driving force in complex II gene history. Although RQ is a metabolic signature for hypoxic metabolism, this metabolite is not necessarily associated with a change in complex I and II composition. Nevertheless, the variety of complex II genes

duplications observed in nematodes and platyhelminths suggests that different adaptations have occurred in different lineages.

Previous studies in *A. suum* have shown that both duplicated complex II subunits (SDHA and SDHD) are exchanged when the environment and oxygen tension change. These adjustments are responsible for the change in the SDH-FRD enzymatic activity and have been associated with the need to use different quinones (Tielens and Van Hellemond, 1998 ; Amino et al., 2000; Hellemond et al., 2003; Iwata et al., 2008; Harada et al., 2013;). In our study, we show that both quinones could bind to both *A. suum* SDHD subunits. Moreover, in organisms that have duplicated the subunits that contact quinones (SDHB, SDHC, and SDHD), no changes are observed in the amino acids that interact directly with quinones. Thus, the switch of subunits does not appear to be related to the quinone-binding capacity.

Gene duplications in complex I have not been previously examined in nematodes and platyhelminths. We searched for the nuclear-encoded quinone-binding subunit genes (*nduf2* and *nduf7*) and found that different gene duplication events occurred among different lineages, similar to the phenomenon observed for complex II subunits. So far, these gene duplications have not been associated with any particular phenotype or adaptation. Residues in NDUF2 and NDUF7 that interact with quinones have been identified in a previous study (Degli Esposti, 2015). This study identified a Tyr residue in *A. suum* NDUF2-1 that differs from Lys or Arg residues found in all other NDUF2 analyzed. We propose that this would be an adaptation of nematodes to RQ binding. Although we found that Tyr is the most frequent amino acid, His and Lys can also be found at this position (**Supplementary Figure 3**).

The preferred oxygen concentration for the free-living nematode *C. elegans* is around 7–14% (Chang and Bargmann, 2008). *C. elegans* possesses complex I and II subunit duplications. To address the role of duplicated complex I and II subunits in this organism, we analyzed their expression pattern during its lifecycle and observed developmental regulation of their expression. A striking observation is that the essential subunits NDUF2-1 and SDHA-1 (those for which KO is lethal) show a similar expression pattern, which is inverted to the expression pattern of the nonessential subunits (NDUF2-2 and SDHA-2). The essential subunits are generally more highly expressed than the nonessential ones, except in the adult worm and early embryo, suggesting that the nonessential subunits are relevant during these stages. Consistent with this interpretation, the *nduf2-2* and *sdha-2* KO strain have reduced brood size compared to the wild-type strain N2. It is worth noting that *nduf2* and *sdha* duplications co-occur in nematodes. Our results suggest that the duplicated subunits could be part of alternative ETC complexes needed to meet different metabolic demands during the lifecycle of the worm and serves to increase the overall organismal fitness. In this sense, it is worth noting that expression of *nduf2-2* and *sdha-2* increase in the early embryo and dauer stage, where a hypoxic metabolism is thought to occur.

The association of an alternative complex II with RQ has been firmly established in *A. suum*, becoming a paradigm in helminth biochemistry. However, our results highlight that *A. suum* is the exception rather than the rule. Furthermore, different variations in the complex I subunits that interact with quinones are also observed in distinct helminth lineages. Dissimilar subunit arrays for ETC complexes appear to have evolved to adapt to different stages of helminth lifecycles, environmental conditions, or to specific cells or tissues within worms.

MATERIAL AND METHODS

Sequences Identification and Analysis

In order to obtain information about helminth complex II and nuclear-encoded quinone-binding complex I subunits, we analyzed genomes and transcriptomes of six platyhelminths (*E. granulosus*, *E. multilocularis*, *H. miscrostoma*, *F. hepatica*, *S. mansoni*, and *S. mediterranea*) and nine nematodes (*T. muris*, *A. suum*, *Brugia malayi*, *Onchocerca volvulus*, *M. hapla*, *S. ratti*, *H. contortus*, *N. americanus*, and *C. elegans*). Sequences were retrieved from https://parasite.wormbase.org (WBPS10), UniProt (https://www.uniprot.org) and *S. mediterranea* database (http://smedgd.neuro.utah.edu).

Searches were performed initially with Protein Basic Local Alignment Search Tool (BLASTP) (protein databases) using *A. suum* and *C. elegans* SDHA, SDHB, SDHC, SDHD, NDUF2, and NDUF7 sequences as queries and confirmed by best reciprocal hits in BLAST. Additionally, TBLASTN searches were performed using genomic sequences and cDNAs databases and the protein sequences previously identified or the most closely related organism's sequences as queries. This served to confirm the annotated protein sequences and to identify the nonannotated ones. Identified sequences were confirmed by best reciprocal hits in BLAST.

To study the evolutionary history of complex II and nuclear-encoded quinone-binding complex I subunits, phylogenetic analysis were performed in the website phylogeny.fr (Dereeper et al., 2008; Dereeper et al., 2010) for every set of identified subunits. Multiple sequence alignments were made with MUSCLE 3.8 (Chojnacki et al., 2017) and then curated with Gblocks 0.91b (Castresana, 2000; Talavera and Castresana, 2007). Phylogeny analyses were made with PhyML 3.0 (Guindon et al., 2010) and tree rending with TreeDyn 198.3 (Chevenet et al., 2006; Dereeper et al., 2008). Alignments for all subunits studied are available in **Supplementary Figure 1**.

Modeling of Complex II Structures

Complex II structures for *A. suum* (normoxia) and *C. elegans* were built by homology modeling with Modeller (Webb and Sali, 2016), using as a 3D template the X-ray structure of *A. suum* (hypoxia) stored in the Protein Data Bank with the code 5C2T. Multiple sequence alignment of complex II between *A. suum* (normoxia and hypoxia) and *C. elegans* was done with Clustal Omega (Chojnacki et al., 2017). To validate the docking technique, we compared the predicted binding sites for RQ_2 with the one from the crystallized structure 5C2T. The best docking solution showed a root mean square deviation of 0.15 nm (**Figure 4C**). A 0.2-nm root mean square deviation cutoff is often used as a criterion of the correct bound structure prediction (Bursulaya et al., 2003).

Docking calculations were performed with AutoDock Vina 1.1.2 (Trott and Olson, 2010). Input PDBQT files for the receptors and ligands were prepared with AutoDockTools 1.5.6 (Morris et al., 2009). All torsions of the ligand were set as fully rotatable, applying to the receptors a partially flexible treatment. The maximum number (10) of rotatable side chains allowed by the program was chosen, selecting residues located closer than 0.6 nm from RQ_2 in 5C2T. List of flexible residues by domain: SDHB: Trp196, Trp197, His240, Ile242; SDHC: Leu60, Trp69, Ser72, Arg76; SDHC: Asp106 and Tyr 107. The search space was defined by a grid with *x*, *y*, and *z* dimensions of 26 × 24 × 24 Å. The grid was centered so that the binding pocket corresponding to RQ_2 could be included in the search space.

Expression Profile Analysis of SDHA and NDUF2 Coding Genes

Transcriptomic data used for this analysis were taken from the GExplore web (Hutter and Suh, 2016).The values of depth of coverage per million bases for each time-point were converted to reads per kilobase per million mapped reads to use a more familiar representation of expression values.

Strains and Culture Conditions

C. elegans wild-type Bristol N2 (N2) and VC393 (*nduf2-2* (*ok437*) III) were provided by the *Caenorhabditis* Genetics Center. VC393 has a 1,448-bp deletion expanding several exons of *nduf2-2*. *C. elegans* TM1420 (*sdha-2* (*tm1420*) I) was obtained from the National BioResource Project. This strain has a 519-bp deletion in exon 4 of *sdha-2*.

Worms were maintained at 20°C on nematode growth medium agar plates seeded with *Escherichia coli* OP50. To obtain synchronized worms, gravid adults on nematode growth medium

agar plates were treated with an alkaline hypochlorite solution, and eggs were collected and placed in M9 medium without food, at 20°C for 24 h (Brenner, 1974).

Brood Size Analysis in *sdha-2 and nduf2-2* KO Strains

Wild-type (N2) and mutant (*sdha-2* KO and *nduf2-2* KO) worms were synchronized, and, after reaching the L4 stage, 15 worms for each strain were separated and plated in groups of 3. When a phenotype was observed, the experiment was performed three times. Offspring production was counted every day for 4 days.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

LO performed the genomic and phylogenetic analyses. CM-R performed the expression analysis and the experiments with *C. elegans* strains. EB and SP performed the molecular modeling studies of complex II with quinones. LO, CM-R, EB, SP, and GS analyzed the data, drafted, and wrote the manuscript. GS conceptualized the study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.01043/full#supplementary-material

REFERENCES

Allen, P. C. (1973). Helminths: comparison of their rhodoquinone. *Exp. Parasitol.* 34, 211–219. doi: 10.1016/0014-4894(73)90080-5

Amino, H., Osanai, A., Miyadera, H., Shinjyo, N., Tomitsuka, E., Taka, H., et al. (2003). Isolation and characterization of the stage-specific cytochrome b small subunit (CybS) of Ascaris suum complex II from the aerobic respiratory chain of larval mitochondria. *Mol. Biochem. Parasitol.* 128, 175–186. doi: 10.1016/S0166-6851(03)00074-4

Amino, H., Wang, H., Hirawake, H., Saruta, F., Mizuchi, D., Mineki, R., et al. (2000). Stage-specific isoforms of Ascaris suum complex. II: The fumarate reductase of the parasitic adult and the succinate dehydrogenase of free-living larvae share a common iron-sulfur subunit. *Mol. Biochem. Parasitol.* 106, 63–76. doi: 10.1016/S0166-6851(99)00200-5

Brenner, S. (1974). The genetics of Caenorhabditis elegans. *Genetics* 77, 71–94

Bursulaya, B. D., Totrov, M., Abagyan, R., and Brooks, C. L. (2003). Comparative study of several algorithms for flexible ligand docking. *J. Comput. Aided. Mol. Des.* 17, 755–763. doi: 10.1023/B:JCAM.0000017496.76572.6f

Cassada, R. C., and Russell, R. L. (1975). The dauer larva, a post-embryonic developmental variant of the nematode Caenorhabditis elegans. *Dev. Biol.* 46, 326–342. doi: 10.1016/0012-1606(75)90109-8

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552. doi: 10.1093/oxfordjournals.molbev.a026334

Chang, A. J., and Bargmann, C. I. (2008). Hypoxia and the HIF-1 transcriptional pathway reorganize a neuronal circuit for oxygen-dependent behavior in Caenorhabditis elegans. *Proc. Natl. Acad. Sci. U.S.A.* 105, 7321–7326. doi: 10.1073/pnas.0802164105

Chevenet, F., Brun, C., Bañuls, A.-L., Jacq, B., and Christen, R. (2006). TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics* 7, 439. doi: 10.1186/1471-2105-7-439

Chojnacki, S., Cowley, A., Lee, J., Foix, A., and Lopez, R. (2017). Programmatic access to bioinformatics tools from EMBL-EBI update: 2017. *Nucleic Acids Res.* 45, W550–W553. doi: 10.1093/nar/gkx273

Degli Esposti, M. (2015). Genome analysis of structure-function relationships in respiratory complex i, an ancient bioenergetic enzyme. *Genome Biol. Evol.* 8, 126–147. doi: 10.1093/gbe/evv239

Dereeper, A., Audic, S., Claverie, J.-M., and Blanc, G. (2010). BLAST-EXPLORER helps you building datasets for phylogenetic analysis. *BMC Evol. Biol.* 10, 8. doi: 10.1186/1471-2148-10-8

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., et al. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* 36, W465–W469. doi: 10.1093/nar/gkn180

Erabi, T., Higuti, T., Kakuno, T., Yamashita, J., Tanaka, M., and Horio, T. (1975). Polarographic studies on ubiquinone-10 and rhodoquinone bound with chromatophores from rhodospirillum rubrum. *J. Biochem.* 78, 795–801. doi: 10.1093/oxfordjournals.jbchem.a130968

Fioravanti, C. F., and Kim, Y. (1988). Rhodoquinone requirement of the Hymenolepis diminuta mitochondrial electron transport system. *Mol. Biochem. Parasit.* 28, 129–134. doi: 10.1016/0166-6851(88)90060-6

Fioravanti, C. F., Walker, D. J., and Sandhu, P. S. (1998). Metabolic transition in the development of Hymenolepis diminuta (Cestoda). *Parasitol. Res.* 84, 777–782. doi: 10.1007/s004360050487

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel,O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010

Harada, S., Inaoka, D. K., Ohmori, J., and Kita, K. (2013). Diversity of parasite complex II. *Biochim. Biophys. Acta - Bioenerg.* 1827, 658–667. doi: 10.1016/j.bbabio.2013.01.005

Hellemond, J. J., van Klei, A., van der Weelden, S. H., and Tielens, A. G. M. (2003). Biochemical and evolutionary aspects of anaerobically functioning mitochondria. *Philos. Trans. R. Soc. London. Ser. B Biol. Sci.* 358, 205–215. doi: 10.1098/rstb.2002.1182

Hutter, H., and Suh, J. (2016). GExplore 1.4: an expanded web interface for queries on Caenorhabditis elegans protein and gene function. *Worm* 5, e1234659. doi: 10.1080/21624054.2016.1234659

Iwata, F., Shinjyo, N., Amino, H., Sakamoto, K., Islam, M. K., Tsuji, N., et al. (2008). Change of subunit composition of mitochondrial complex II (succinate–ubiquinone reductase/quinol–fumarate reductase) in Ascaris suum during the migration in the experimental host. *Parasitol. Int.* 57, 54–61. doi: 10.1016/j.parint.2007.08.002

Kuramochi, T., Hirawake, H., Kojima, S., Takamiya, S., Furushima, R., Aoki, T., et al. (1994). Sequence comparison between the flavoprotein subunit of the fumarate reductase (Complex II) of the anaerobic parasitic nematode, Ascaris suum and the succinate dehydrogenase of the aerobic, free-living nematode, Caenorhabditis elegans. *Mol. Biochem. Parasitol.* 68, 177–187. doi: 10.1016/0166-6851(94)90163-5

Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* doi: 10.1093/nar/gkz239

Matsumoto, J., Sakamoto, K., Shinjyo, N., Kido, Y., Yamamoto, N., Yagi, K., et al. (2008). Anaerobic NADH-Fumarate reductase system is predominant in the respiratory chain of echinococcus multilocularis, providing a novel target for the chemotherapy of alveolar echinococcosis. *Antimicrob. Agents Chemother.* 52, 164–170. doi: 10.1128/AAC.00378-07

Moore, H. W., and Folkers, K. (1965). Coenzyme Q. LXII. Structure and synthesis of rhodoquinone, a natural aminoquinone of the coenzyme q group 1. *J. Am. Chem. Soc.* 87, 1409–1410. doi: 10.1021/ja01084a065

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., et al. (2009). AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* 30, 2785–2791. doi: 10.1002/jcc.21256

Roos, M. H., and Tielens, A. G. (1994). Differential expression of two succinate dehydrogenase subunit-B genes and a transition in energy metabolism during the development of the parasitic nematode Haemonchus contortus. *Mol. Biochem. Parasitol.* 66, 273–281. doi: 10.1016/0166-6851(94)90154-6

Saruta, F., Kuramochi, T., Nakamura, K., Takamiya, S., Yu, Y., Aoki, T., et al. (1995). Stage-specific isoforms of complex II (succinate-ubiquinone oxidoreductase) in mitochondria from the parasitic nematode, Ascaris suum. *J. Biol. Chem.* 270, 928–932. doi: 10.1074/jbc.270.2.928

Sato, M., and Ozawa, H. (1969). Occurrence of ubiquinone and rhodoquinone in parasitic nematodes, Metastrangylus elongatus and Ascaris lumbricoides var. suis. *J. Biochem.* 65, 861–867. doi: 10.1093/oxfordjournals.jbchem.a129090.

Shiobara, Y., Harada, C., Shiota, T., Sakamoto, K., Kita, K., Tanaka, S., et al. (2015). Knockdown of the coenzyme Q synthesis gene Smed-dlp1 affects planarian

regeneration and tissue homeostasis. *Redox Biol.* 6, 599–606. doi: 10.1016/j.redox.2015.10.004

Takamiya, S., Kita, K., Wang, H., Weinstein, P. P., Hiraishi, A., Oya, H., et al. (1993). Developmental changes in the respiratory chain of Ascaris mitochondria. *Biochim. Biophys. Acta - Bioenerg.* 1141, 65–74. doi: 10.1016/0005-2728(93)90190-Q

Takamiya, S., Matsui, T., Taka, H., Murayama, K., Matsuda, M., and Aoki, T. (1999). Free-living nematodes Caenorhabditis elegans possess in their mitochondria an additional rhodoquinone, an essential component of the eukaryotic fumarate reductase system. *Arch. Biochem. Biophys.* 371, 284–289. doi: 10.1006/abbi.1999.1465

Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* 56, 564–577. doi: 10.1080/10635150701472164

Tielens, A. G. M., and Van Hellemond, J. J. (1998). The electron transport chain in anaerobically functioning eukaryotes. *Biochim. Biophys. Acta - Bioenerg.* 1365, 71–78. doi: 10.1016/S0005-2728(98)00045-0

Trott, O., and Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31, 455–461. doi: 10.1002/jcc.21334

Van Hellemond, J. J., Klockiewicz, M., Gaasenbeek, C. P., Roos, M. H., and Tielens, A. G. (1995). Rhodoquinone and complex II of the electron transport chain in anaerobically functioning eukaryotes. *J. Biol. Chem.* 270, 31065–31070. doi: 10.1074/jbc.270.52.31065

Van Hellemond, J. J., Van Remoortere, A., and Tielens, A. G. (1997). Schistosoma mansoni sporocysts contain rhodoquinone and produce succinate by fumarate reduction. *Parasitol.* 115 (Pt 2), 177–182. doi: 10.1017/S003118209700125X

Webb, B., and Sali, A. (2016). "Comparative Protein Structure Modeling Using MODELLER," in *Current Protocols in Protein Science* (Hoboken, NJ, USA: John Wiley & Sons, Inc.), 2.9.1–2.9.37. doi: 10.1002/cpps.20

The Effect of Gut Microbiome Composition on Human Immune Responses: An Exploration of Interference by Helminth Infections

Ivonne Martin[1,2]*[†], Maria M. M. Kaisar[3]†, Aprilianto E. Wiria[3], Firdaus Hamid[4], Yenny Djuardi[3], Erliyani Sartono[5], Bruce A. Rosa[6], Makedonka Mitreva[6,7], Taniawati Supali[3], Jeanine J. Houwing-Duistermaat[8,9,10]‡, Maria Yazdanbakhsh[5]‡ and Linda J. Wammes[5]

[1] Department of Mathematics, Faculty of Information Technology and Science, Parahyangan Catholic University, Bandung, Indonesia, [2] Department of Biomedical Data Sciences, section Medical Statistics, Leiden University Medical Center, Leiden, Netherlands, [3] Department of Parasitology, Faculty of Medicine, Universitas Indonesia, Depok, Indonesia, [4] Department of Microbiology, Faculty of Medicine, Universitas Hasanuddin, Makassar, Indonesia, [5] Department of Parasitology, Leiden University Medical Center, Leiden, Netherlands, [6] McDonnell Genome Institute, Washington University in St. Louis, St. Louis, MO, United States, [7] Department of Medicine, Washington University School of Medicine, St. Louis, MO, United States, [8] Department of Statistics, University of Leeds, Leeds, United Kingdom, [9] Department of Biostatistics and Research Support, Julius Center, University Medical Centre Utrecht, Utrecht, Netherlands, [10] Alan Turing Institute, London, United Kingdom

Background: Soil-transmitted helminths have been shown to have the immune regulatory capacity, which they use to enhance their long term survival within their host. As these parasites reside in the gastrointestinal tract, they might modulate the immune system through altering the gut bacterial composition. Although the relationships between helminth infections or the microbiome with the immune system have been studied separately, their combined interactions are largely unknown. In this study we aim to analyze the relationship between bacterial communities with cytokine response in the presence or absence of helminth infections.

Results: For 66 subjects from a randomized placebo-controlled trial, stool and blood samples were available at both baseline and 21 months after starting three-monthly albendazole treatment. The stool samples were used to identify the helminth infection status and fecal microbiota composition, while whole blood samples were cultured to obtain cytokine responses to innate and adaptive stimuli. When subjects were free of helminth infection (helminth-negative), increasing proportions of *Bacteroidetes* was associated with lower levels of IL-10 response to LPS {estimate [95% confidence interval (CI)] −1.96 (−3.05, −0.87)}. This association was significantly diminished when subjects were helminth-infected (helminth positive) (*p*-value for the difference between helminth-negative versus helminth-positive was 0.002). Higher diversity was associated with greater IFN-γ responses to PHA in helminth-negative (0.95 (0.15, 1.75); versus helminth-positive [−0.07 (−0.88, 0.73), *p*-value = 0.056] subjects. Albendazole treatment showed no direct effect in the association between bacterial proportion and cytokine responses, although the *Bacteroidetes'* effect on IL-10 responses to LPS tended downward in the albendazole-treated group [−1.74 (−4.08, 0.59)] versus placebo [−0.11 (−0.84, 0.62); *p*-value = 0.193].

Conclusion: We observed differences in the relationship between gut microbiome composition and immune responses, when comparing individuals infected or uninfected with geohelminths. Although these findings are part of a preliminary exploration, the data support the hypothesis that intestinal helminths may modulate immune responses, in unison with the gut microbiota.

Trial Registration: ISRCTN, ISRCTN83830814. Registered 27 February 2008 — Retrospectively registered, http://www.isrctn.com/ISRCTN83830814.

Keywords: helminth, gut microbiome, whole blood cytokine, interleukin-10, *Bacteroidetes*, diversity, randomized controlled trial

INTRODUCTION

Diseases of modernity, such as allergy, autoinflammatory, and metabolic diseases are increasingly observed in industrialized countries. It has been speculated that this growing rate was caused by changes in lifestyle, diet and environmental factors, such as pollutant exposure or hygiene. Hygiene improvement has dramatically decreased the prevalence of certain infectious agents such as parasitic helminths while these may have protective effects against autoinflammatory diseases (Wammes et al., 2016). Studies analyzing the capacity of helminths to modulate the immune system have been carried out in recent decades. However, it has become clear that this is an interplay with several other factors, such as diet, environment and also other gut inhabitants, such as the microbiota.

Early studies showed that gut microbiota is involved in developmental aspects of the immune system and that disturbance can lead to autoinflammatory disorders (Round and Mazmanian, 2009). Already in 1963, it was reported that the immune system of germ-free mice failed to respond to molecular patterns of pathogenic and beneficial microorganisms, causing morphological tissue defects in the intestinal wall (Abrams et al., 1963). In healthy humans, the role of gut microbiota and immune response was studied more recently. It was found that certain bacteria are beneficial for development and function of the immune system and simultaneously the immune system can influence the composition or function of gut microbiota, all relating to inflammatory disorders [reviewed in Belkaid and Hand (2014)].

The presence of parasitic helminths in the gastrointestinal tract may exert a direct influence on the host's gut microbiome as they share the same niche. Although in animal models helminths were shown to increase microbial abundance and diversity (Reynolds et al., 2015), the findings in human studies are not consistent. Several studies analyzing the effect of helminth on gut microbiota have indicated a higher diversity of gut microbiota in helminth-positive subjects compared to helminth-negative subjects (Lee et al., 2014; Ramanan et al., 2016). A study in Ecuador showed that this difference in diversity might be related to specific helminth species since they did not find any alterations in *Trichuris trichiura*-infected children (Cooper et al., 2013). This might be influenced by different factors among which are different bacterial profiling techniques or confounders such as ethnicity, anthelminthic treatment, and environmental differences.

As it has been shown that changes in both gut microbiota and helminth infection status might affect the host's immune response, it is suspected that the presence of helminth might directly or indirectly affect the immune system by altering the gut microbial community (Zaiss et al., 2015). For instance, the transfer of the microbiota of *Heligmosomoides polygyrus bakeri*-infected mice to uninfected mice induced similar protection against allergic airway inflammation as observed with helminth infection (Wilson et al., 2005). In humans, studies on the triangular relation between helminth with the microbiome and immune system are still in infancy.

To our knowledge, the number of longitudinal studies analyzing the association between gut microbiota and immune responses in helminth-endemic areas is still limited. To understand the interaction of the gut microbial community and helminths and their common effect on immune responses, we used data from a household cluster-randomized, double blind, placebo-controlled trial of albendazole treatment in a helminth-endemic area. In this study, it has been shown that deworming reduced helminth prevalence and consequently increased several whole blood cytokine responses (Wammes et al., 2016). Helminth infection and anthelminthic treatment separately did not change the gut microbiota (Martin et al., 2018). However, when subjects remained infected while treated with albendazole, a decrease of *Bacteroidetes*: *Firmicutes* ratio and an increase of *Actinobacteria*: *Firmicutes* ratio were observed, leading to the hypothesis that there is a cross-talk between microbiome composition and immune response which is disrupted by the presence of helminths and that removing helminth by anthelminthic might affect this communication. Our aim was to characterize the association between bacterial relative abundance with the whole blood cytokine responses and the effect of helminth infections and deworming on this interaction.

MATERIALS AND METHODS

Participants

Stool samples from 150 subjects from the immunoSPIN study (Wiria et al., 2010) taking place in Ende subdistrict, Indonesia, were analyzed for the fecal microbiome. From these 150 subjects, 66 subjects were included in this study based on the complete stool data and available cytokine measurements before and 21 months after the first treatment. The microbiome composition

---

Abbreviations: AscAg, *Ascaris* antigen; BMI, Body mass index; IFN, interferon; IL, interleukin; LPS, lipoposaccharide; PHA, Phytohaemagglutinin; TNF, Tumor necrosis factor.

for these subjects are representative of these 150 subjects (**Table S1**). Four different helminth species were found namely *Ascaris lumbricoides*, hookworms (*Necator americanus* and *Ancylostoma duodenale*) and *T. trichiura*. Details on sample collection and measuring the infection status using PCR are described elsewhere (Wiria et al., 2010). *T. trichiura* infection was assessed only by microscopy, since at that time there was no real-time PCR data available for this species. For this manuscript, we defined a helminth-infected subject as a participant with a positive real-time PCR (cycle of threshold (Ct) value ≤ 30) and/or positive microscopy for one or more species of helminths, as described previously (Martin et al., 2018). Subjects with a positive real-time PCR with a Ct above 30 were regarded as uninfected.

Microbiome Composition

The amplification and pyrosequencing of the 16S rRNA gene followed the protocols developed by the Human Microbiome Project (HMP) (A framework for human microbiome research. Nature, 2012) at the McDonnell Genome Institute, Washington University School of Medicine in St. Louis and have been described previously elsewhere (Martin et al., 2018; Rosa et al., 2018). Briefly, the V1–V3 hypervariable region was PCR — amplified and the PCR products were sequenced on the Genome Sequencer Titanium FLX (Roche Diagnostics, Indianapolis, Indiana), generating on average 6,000 reads per sample. Details of the filtering and analytical processing of 16S rRNA data for this cohort has been previously described in Rosa et al. (2018). The assembled contigs count data as a result of RDP classification was organized in a matrix format with taxa in columns and subjects in rows. The entries in the table represent the number of reads for each taxon for each subject. Our work is focused at a phylum level of gut bacteria. Five bacterial phyla have average relative abundances larger than 1%, namely *Actinobacteria*, *Bacteroidetes*, *Firmicutes*, *Proteobacteria* and an unclassified category, which consists of sequences which could not be categorized into a phylum. The remaining bacterial phyla which had lower relative abundance were pooled together into a pooled category. In the analysis, we retained the count for the three most abundant bacterial phylum proportions, namely *Actinobacteria*, *Bacteroidetes*, and *Firmicutes*. The proportion for each phylum was obtained by dividing each sequence count by the total sequence per person at each time point. Along with bacterial proportions, we computed at a phylum level the bacterial diversity within samples (Shannon index) and between samples (Bray-Curtis dissimilarity) using R package vegan (Oksanen et al., 2017). The Shannon index represents not only the presence of taxa but also the abundance of corresponding taxa. The higher diversity index means that there was not a single taxon dominating the community and the total bacterial abundance is spread out over all taxa. The Bray Curtis dissimilarity measures the percentage of similarity between one sample from the other with values range from 0 (completely similar) to 1 (completely dissimilar).

Whole Blood Cytokine Responses

The method to obtain and assess the cytokines responses were described elsewhere (Wiria et al., 2010). In brief, heparinized blood was diluted 1:4 and cultured in 96-well plates. Plates were incubated for 24 (innate responses) or 72 (adaptive responses) hours, after which supernatants were harvested and stored in freezers. Cytokine levels were measured by Luminex bead technology in samples obtained at before and 21 months after the start of treatment.

The analyses carried in this manuscript are limited to innate responses [interleukin (IL)-10 and tumor necrosis factor-alpha (TNF-α)] to lipopolysaccharide (LPS) from *E. coli* and adaptive responses [interferon-gamma (IFN-γ) and IL-5) to *Ascaris* antigen (AscAg) and general T cell stimulator phytohemagglutinin (PHA)]. The AscAg was a homogenate of adult worm *A. lumbricoides* obtained from infected human (Wammes et al., 2014).

Statistical Method

All parameters of interest were described as means or frequency (± standard deviation). Prevalence rates were calculated and compared using the Pearson chi-square test, while the Student *t*-test was used to compare continuous variables.

To study the relationship between cytokines and microbiome over the two time points, a linear mixed effect regression model was fitted with helminth status and treatment as covariates. All models have been adjusted with age and sex, however, since both covariates were not significantly associated with the cytokine responses in any model, they are not included in the final analysis. The correlation between observations from the same subjects was modeled by including a subject-specific random effect. The microbiome was included in the model either as a bacterial proportion or by the Shannon diversity index. The cytokine responses were $\log_{10}$-transformed [$\log_{10}$(concentration + 1)] to obtain normally distributed variables. First, we analyzed the main effect of bacterial proportion and diversity on cytokine responses. Second, to allow for different effect sizes of bacterial proportion or diversity on cytokine responses in helminth-positive versus -negative subjects, an interaction term between bacterial categories and infection was included in the model. The *p*-value for this interaction term indicated the statistical evidence for different effect sizes in helminth-positive or -negative groups. Due to limited sample size, we restricted the analysis into estimating the general effect of helminth infection on the relationship between bacterial proportion and cytokine responses. In the same manner, due to the limitation of sample size, the direct treatment effect cannot be estimated. This means that the estimates for the infection effect on the relationship between bacterial proportions and diversity on cytokine responses were obtained from data of subjects who were infected at any time point, regardless of their randomization arm. However, as treatment removes helminth infection, the analysis of treatment effect on the relationship between bacterial proportion and cytokine responses is viewed as a proxy to understand the role of helminth in this relationship.

For this purpose, a linear mixed effect model was fitted with bacterial proportion or diversity, treatment, and time as covariates. This model was able to characterize three different associations, namely the association between bacterial proportion or diversity

on cytokines at pre-treatment, the difference of the association at pre-treatment, and at post-treatment in the placebo group (time effect), as well as the difference of the association at post-treatment between albendazole and placebo group.

For each outcome separately, these models were fitted on subjects who at least had an observation at pre-treatment. The lme4 package in statistical software R was used for model fitting. The significance of the covariate effect was obtained from the likelihood ratio test. Bonferroni correction was used to adjust for multiple testing. We have applied Bonferroni correction for the number of non-correlated cytokine per stimulatory condition resulting in dividing the alpha cut-off level for significance by 2 (LPS responses) and 3(AscAg responses) and 2(PHA responses). The statistical analyses were performed in R (R Core Team, 2017) with mainly lme4 and lmerTest packages (Bates et al., 2015; Kuznetsova et al., 2017). The full record was created using the knitr package in R (Xie, 2017) and is available online at https://github.com/Helminths_GutMicrobes_Cytokine/.

RESULTS

The Effect of Bacterial Proportions and Diversity on *In Vitro* Cytokine Responses

Since it is hypothesized that gut bacteria are associated with certain cytokine responses and thereby possibly immune disorders, we set out to explore this relationship by using data from the ImmunoSPIN trial. For 66 subjects, cytokine responses were measured at pre-treatment and 21 months after the start of anthelmintic treatment. At baseline, 40 out of 66 (60.6%) individuals in Ende were infected with one or more helminth species, and hookworm was the most dominant species (31.8%) followed by *A. lumbricoides* (25.7%) and *T. trichiura* (22.7%). The baseline characteristics such as age, gender, BMI and helminth prevalence were similar between the two treatment arms (**Table 1**). Three-monthly albendazole treatment for 21 months reduced the infection prevalence from 65.4% to 19.2% versus a slight increased of helminth infections from 57.5% to 65% in the placebo group (**Table S2**).

We analyzed proportions of three bacterial phyla (*Actinobacteria*, *Bacteroidetes*, and *Firmicutes*) as these were most abundant in our study population. For the cytokine responses, we selected outcomes representative of different parts of the immune system. We have opted for the pro- and anti- inflammatory (TNF and IL-10 respectively) immune response to LPS to represent the innate responses, and the Th1 and Th2 signature cytokines (IFN-gamma and IL-5 respectively) for the adaptive response to a helminth antigen (AscAg) and T-cell stimulatory PHA. When fitting the linear mixed model, no direct effect was observed of bacterial proportions or Shannon diversity on whole blood cytokine responses (**Table 2**).

Interference by Helminth Infection in the Effect of Bacterial Proportions and Diversity on *In Vitro* Cytokine Responses

To elucidate the possible role of helminth infections in the interplay of bacteria and immune responses, we conducted analyses in helminth-positive and -negative groups. For this

purpose, we used observations at both pre-treatment and post-treatment. Regardless of the randomization arm, we fitted the linear mixed model on each cytokine responses as outcomes. The predictors were bacterial proportions and its interaction with helminth infection. A similar analysis was performed to estimate the association between bacterial diversity and cytokine responses. **Figure 1** illustrates the associations between bacterial proportions or diversity and cytokine responses when subjects were helminth-positive or -negative.

In the innate immune response to LPS, the *Bacteroidetes* proportion showed a significant negative association with IL-10 levels in helminth-negative subjects {estimated effect [95% confidence interval (CI)]: −1.96 (−3.05, −0.87), p-value = 0.001; **Figure 1B**} which shows that a unit increase of *Bacteroidetes* proportion will reduce the concentration of IL-10 to LPS almost twice as much. This association was significantly different from that of helminth-negative subjects (p-value for the difference = 0.002, **Figure 1**) in which the association was absent [−0.03 (−0.59, 0.53)]. The bacterial diversity had no significant association with IL-10 response to LPS (**Figure 1**). With regard to the helminth-specific cytokine responses, none of IFN-γ and IL-5 responses to AscAg were significantly associated with bacterial proportions or diversity (**Figure 1**). In the adaptive responses (PHA), none of the cytokine responses were significantly associated with the bacterial proportion in uninfected subjects (**Figure 1**). Although not significant, we noticed lower levels of IFN-γ to PHA with higher *Firmicutes* proportions [−1.57 (−3.08, −0.05), p-value = 0.045; **Figure 1**]. This association between *Firmicutes* proportion with IFN-γ to PHA in uninfected subjects was however significantly different from that in subjects who were infected (p-value for the difference = 0.009, **Figure 1**). At the same time, there was a significantly increasing concentration of IFN-γ to PHA among those who were uninfected when bacterial diversity was higher [0.95 (0.15, 1.75), p-value 0.022; **Figure 1**], although this association was not significantly different from the helminth-positive group [−0.07 (−0.88, 0.73), p-value for the difference = 0.056; **Figure 1**]. A similar negative association of *Firmicutes* was observed in IL-5 responses to PHA in uninfected subjects [−1.52 (−3.02, −0.02), p-value = 0.05; **Figure 1**]. Conversely increasing bacterial diversity led to slightly higher levels of IL-5 to PHA in the uninfected subjects [0.85 (0.07, 1.63), p-value = 0.034; **Figure 1**]. Both observations were not significantly different from the effects in those who were helminth-positive.

The Effect of Albendazole on the Relationship Between Bacterial Proportion and Diversity and *In Vitro* Cytokine Responses

We further investigated whether deworming affects the relationship between bacterial proportions or diversity and cytokine responses. For this purpose, we fitted the linear mixed model on all subjects (n = 66) to characterize the association between bacterial proportions and cytokine responses at two time points and in the two randomization arms. These analyses were irrespective of the infection status. A similar model was applied for the diversity index.

TABLE 1 | Characteristics of the participants at baseline.

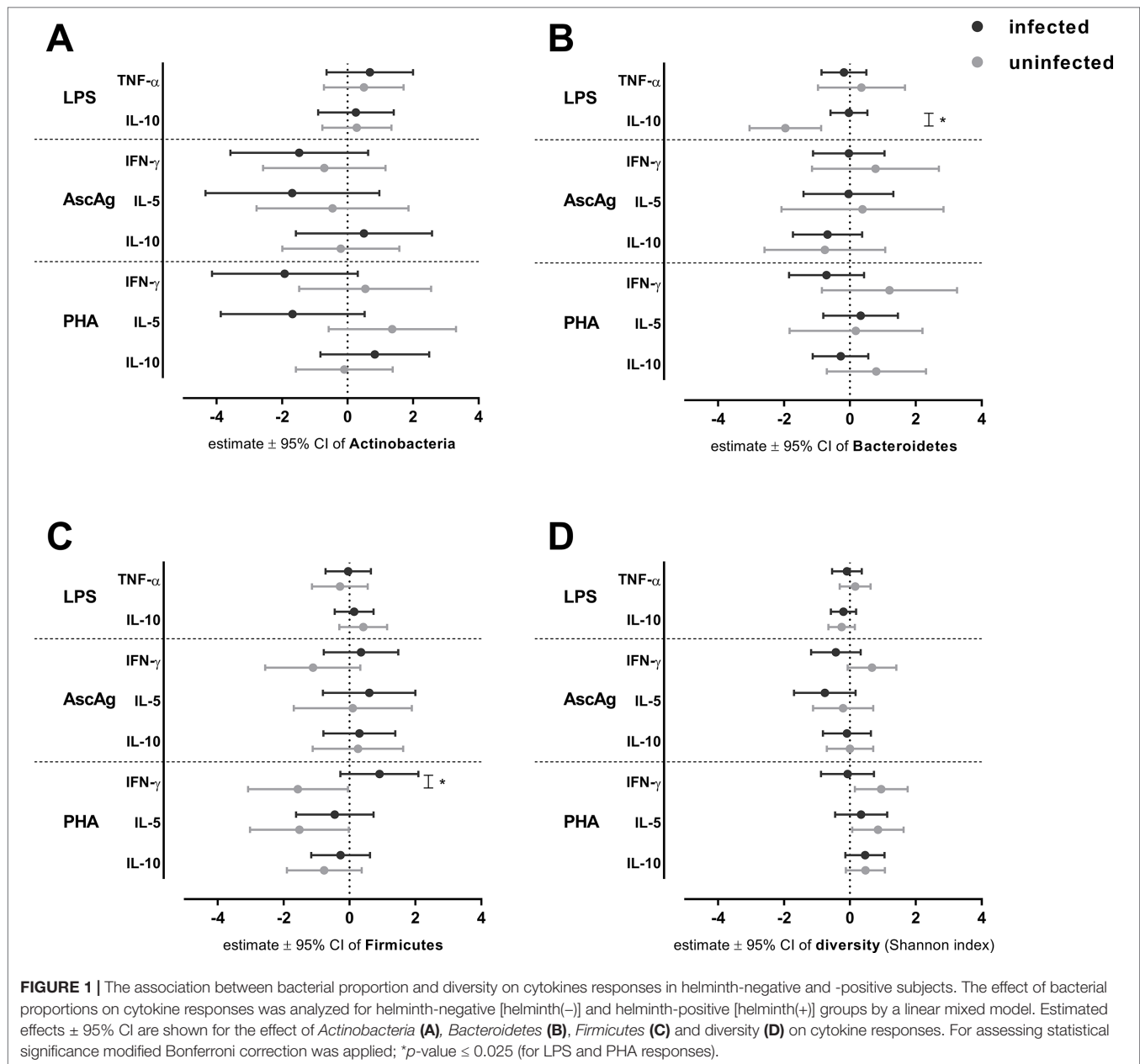| Characteristics	Albendazole		Placebo	
	N	Result	N	Result
Gender, female [N (%)]	26	12 (46.1)	40	22 (55.0)
Age [mean (SD)]	26	27.3 (16.1)	40	26.7 (15.7)
Children [< = 18 years old; N (%)]		10 (38.4)		17 (42.5.0)
Adults [>18 years old; N (%)]		16 (61.5)		23 (57.5)
zBMI [mean (SD)]	10	−0.52 (0.98)	17	−0.83 (0.64)
BMI [mean (SD)]	16	23.39 (3.44)	23	23.49 (4.89)
Parasite infection [N (%)]				
A. lumbricoides[a]	26	9 (34.6)	40	8 (20.0)
Hookworm	26	11 (42.3)	40	10 (25.0)
N. americanus[a]	26	10 (38.5)	40	10 (25.0)
A. duodenale[a]	26	2 (7.7)	40	2 (5.0)
T. trichiura[b]	26	5 (19.2)	40	10 (25.0)
Any helminth	26	17 (65.4)	40	23 (57.5)
Abundance of bacterial phyla [mean % (SD)]	26		40	
Actinobacteria		9.1 (5.8)		9.6 (7.7)
Bacteroidetes		8.0 (10.5)		7.1 (12.1)
Firmicutes		72.4 (10.4)		71.6 (12.8)
Proteobacteria		8.5 (7.3)		8.6 (6.2)
unclassified bacteria[#]		1.4 (0.8)		2.5 (3.1)
pooled[*]		0.6 (0.8)		0.6 (0.9)
Diversity Index, median (IQR)	26		40	
Shannon index		0.85 (0.71, 0.99)		0.84 (0.73, 1.00)
Bray-Curtis		0.19 (0.12, 0.26)		0.19 (0.13, 0.28)
Cytokine responses (pg/mL, median, IQR)				
LPS TNF-α	25	664.00 (294.00,1029.00)	40	550.50 (343.00, 840.00)
IL-10	25	242.00 (132.00, 400.00)	40	213.50 (142.00, 380.20)
AscAg IFN-γ	23	28.50 (12.10, 111.80)	37	17.40 (7.74, 60.90)
IL-5	22	32.55 (9.55, 58.42)	37	18.90 (12.00, 62.00)
IL-10	23	8.28 (3.22, 22.45)	37	6.61 (2.55, 10.30)
PHA IFN-γ	23	2,449.00 (354.00, 5,424.00)	37	2299.00 (997.00, 3829.00)
IL-5	23	490.00 (276.00, 747.50)	37	515.00 (333.00, 870.00)
IL-10	23	101.00 (54.05, 167.00)	37	83.20 (36.80, 140.00)

[a]Diagnosed by real-time PCR; [b]diagnosed by microscopy; [#]unclassified bacteria represents the category of sequences that could not be assigned to a phyla, and the; [*]pooled category consists of the remaining 13 phyla with average relative abundance less than 1%.

TABLE 2 | The association between bacterial proportion and diversity on cytokine responses.

		Estimated effect (95% CI)			
		Actinobacteria	*Bacteroidetes*	*Firmicutes*	Shannon
LPS	**IL-10**	0.20 (−0.58, 0.98)	−0.39 (−0.90, 0.12)	0.24 (−0.23, 0.71)	−0.22 (−0.51, 0.07)
	TFN-α	0.55 (−0.35, 1.44)	−0.06 (−0.66, 0.54)	−0.14 (−0.70, 0.41)	0.03 (−0.31, 0.37)
AscAg	**IFN-γ**	−1.03 (−2.45, 0.39)	0.15 (−0.80, 1.10)	−0.20 (−1.13, 0.74)	0.14 (−0.44, 0.71)
	IL-5	−1.02 (−2.78, 0.74)	0.09 (−1.10, 1.28)	0.39 (−0.74, 1.52)	−0.48 (−1.16, 0.20)
	IL-10	−0.04 (−0.56, 0.48)	0.15 (−1.21, 1.51)	−0.72 (−1.63, 0.18)	0.30 (−0.57, 1.16)
PHA	**IFN-γ**	−0.57 (−2.12, 0.98)	−0.26 (−1.28, 0.75)	−0.03 (−1.05, 0.99)	0.45 (−0.18, 1.08)
	IL-5	−0.04 (−1.55, 1.46)	0.32 (−0.67, 1.32)	−0.85 (−1.82, 0.11)	0.61 (0.02, 1.20)
	IL-10	0.46 (0.02, 0.91)	0.33 (−0.80, 1.46)	−0.03 (−0.78, 0.72)	−0.46 (−1.19, 0.27)

Figure 2 displays the associations between the proportions of three major bacterial phyla and diversity with cytokine responses, before and after anthelminthic treatment. With regard to the relationship between *Bacteroidetes* and IL-10 response to LPS, no significant differences were observed between pre- versus post-treatment or between treatment groups (**Figure 2**). While the estimated association between *Bacteroidetes* proportion and IL-10 to LPS at pre-treatment [estimate (95% CI): −0.47 (−1.23, 0.29)] and post-treatment in placebo group [−0.11 (−0.84, 0.62)] were close to zero, the association at post-treatment in
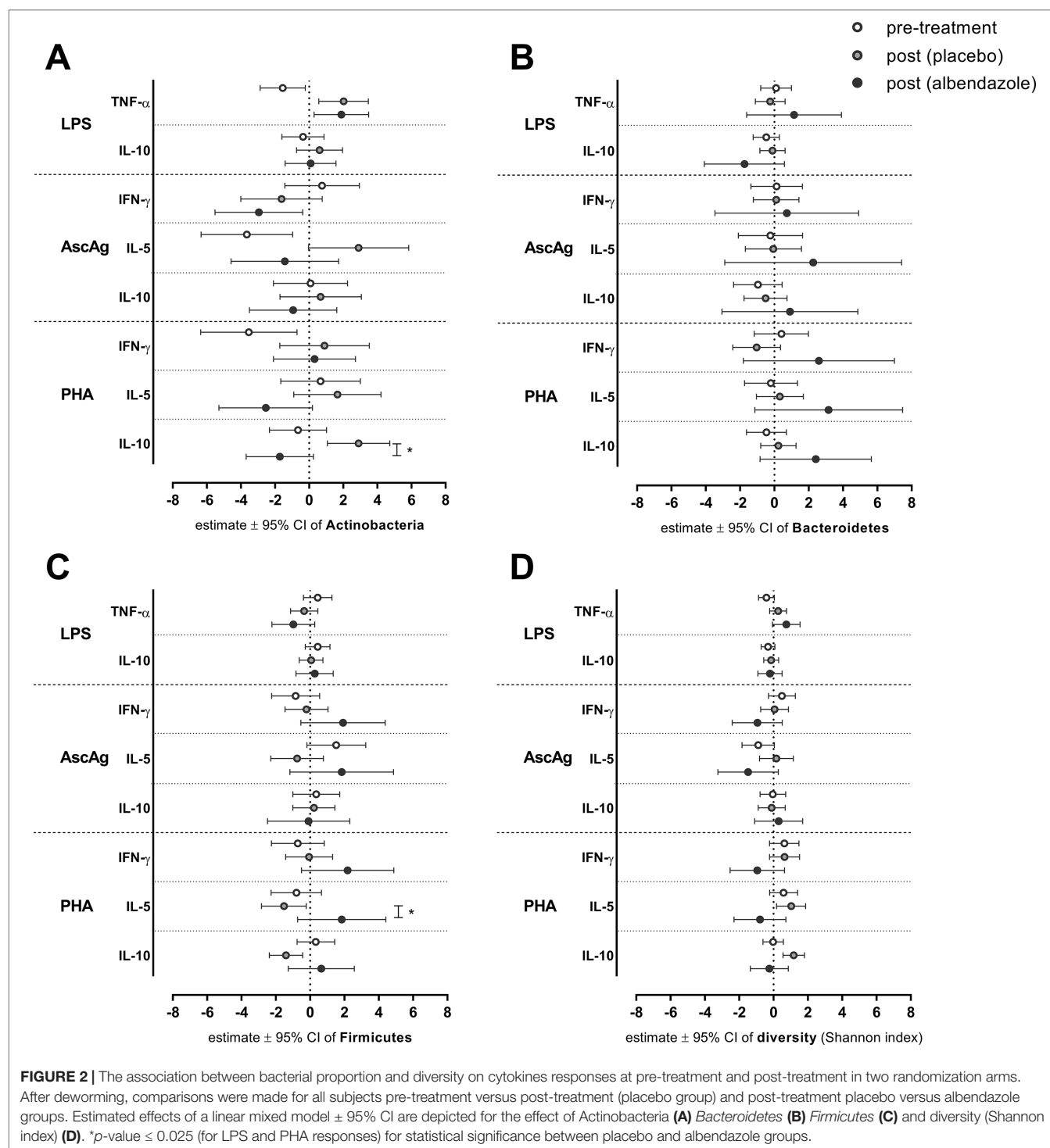
FIGURE 1 | The association between bacterial proportion and diversity on cytokines responses in helminth-negative and -positive subjects. The effect of bacterial proportions on cytokine responses was analyzed for helminth-negative [helminth(−)] and helminth-positive [helminth(+)] groups by a linear mixed model. Estimated effects ± 95% CI are shown for the effect of *Actinobacteria* **(A)**, *Bacteroidetes* **(B)**, *Firmicutes* **(C)** and diversity **(D)** on cytokine responses. For assessing statistical significance modified Bonferroni correction was applied; *p-value ≤ 0.025 (for LPS and PHA responses).

albendazole group was clearly lower [−1.74 (−4.08, 0.59); *p*-value for the difference between placebo and albendazole was 0.193, **Figure 2B**]. The association between IFN-γ in response to PHA and bacterial diversity was also not significant at post-treatment either in the placebo or in the albendazole group (**Figure 2D**).

The association between higher *Actinobacteria* proportion with decreasing response of TNF-α to LPS was borderline significant at pre-treatment [estimate (95% CI): −1.55 (−2.87, −0.22), *p*-value = 0.024; **Figure 2A**]. This association was significantly different to the effect of *Actinobacteria* at post-treatment when subject received placebo [2.02 (0.57, 3.47); *p*-value < 0.001; **Figure 2A**], however no difference was observed when comparing placebo and albendazole group [1.89 (0.29, 3.49), *p*-value for the difference = 0.907; **Figure 2A**]. A similar result was obtained from

the association between *Actinobacteria* with IL-5 responses to AscAg. At pre-treatment, the increasing *Actinobacteria* proportions were significantly associated with less IL-5 production in response to AscAg [−3.65: (−6.34, −0.97), *p*-value = 0.009; **Figure 2A**]. This association was significantly different to the effect of *Actinobacteria* at post-treatment in placebo group [2.90 (−0.03, 5.84), *p*-value = 0.002]. Although the estimated association in albendazole group was lower [−1.42 (−4.58, 1.73), this was not significantly different between the treatment groups (*p*-value = 0.052)]. The effect of Actinobacteria on IL-10 responses to PHA changed after treatment [−0.65 (−2.33–1.02) vs 2.90 (1.07–4.73) for pre-treatment vs post-treatment placebo group, *p*-value 0.005] and was different between placebo and albendazole groups as well [albendazole group −1.72 (−3.69–0.25), *p*-value for placebo vs albendazole 0.001; **Figure 2A**].

FIGURE 2 | The association between bacterial proportion and diversity on cytokines responses at pre-treatment and post-treatment in two randomization arms. After deworming, comparisons were made for all subjects pre-treatment versus post-treatment (placebo group) and post-treatment placebo versus albendazole groups. Estimated effects of a linear mixed model ± 95% CI are depicted for the effect of Actinobacteria **(A)** *Bacteroidetes* **(B)** *Firmicutes* **(C)** and diversity (Shannon index) **(D)**. *p-value ≤ 0.025 (for LPS and PHA responses) for statistical significance between placebo and albendazole groups.

Moreover, while the association between *Firmicutes* and IL-5 response to PHA at pre-treatment was not significantly different compared to the association at post-treatment in placebo group, there was a significant difference of this association between albendazole and placebo group at post-treatment [estimate (95% CI) for placebo −1.52 (−2.83, −0.22) versus albendazole 1.84 (−0.73, 4.41), *p*-value = 0.024; **Figure 2C**].

DISCUSSION

This study aimed to analyze the effect of helminth infections on the relationship between gut microbiota and the immune system. We found a negative association between proportions of *Bacteroidetes* and IL-10 response to LPS in helminth-negative subjects and the presence of helminths was shown to dampen

this effect. Anthelminthic treatment partly recovered this effect, although not statistically significant. To our knowledge, this is the first time that the association between the gut microbiome, presence of parasitic helminths and whole blood cytokine responses was analyzed in a longitudinal study using a randomized placebo-controlled anthelminthic trial.

IL-10 was already marked as a key anti-inflammatory cytokine involved in the induction of immune suppression by helminths (Yazdanbakhsh et al., 2002). Our observation that helminths counteract the suppressed IL-10 response to LPS in subjects with higher *Bacteroidetes* proportions supports the so called "old friends hypothesis" (Rook, 2009), stating that certain infectious agents such as helminths may have protective effects against immune dysfunction and inflammatory diseases, possibly through IL-10. This is strengthened by our observed gradient of the relative abundance of *Bacteroidetes* from rural to urban areas, where immune-related diseases are more prevalent (Bach, 2002). In contrast, a recent meta-analysis indicated that inflammatory bowel disease (IBD) patients displayed lower proportions of *Bacteroidetes* (Zhou and Zhi, 2016), however, this was only found when measuring by real-time quantitative PCR (not by conventional culture) and mainly in Asian studies. Furthermore, a member of the *Bacteroidetes* family, gut inhabitant *Bacteroides fragilis*, was shown to protect mice from experimental colitis, mediated by polysaccharide A (PSA) possibly through IL-10 induction (Mazmanian et al., 2008). However, although *B. fragilis* is the most well-known pathogen of the *Bacteroidetes*, it is the least common species in the *Bacteroidetes* phylum in the human gut (Wexler, 2007). It could therefore be that other factors or species play a dominant role in the general effect of *Bacteroidetes* on IL-10 responses. Further studies are therefore needed to assess the translation of our findings to a clinical setting, for example prevalence or activity of IBD or other auto-immune diseases. Moreover, since we have measured systemic whole blood cytokine responses, we are not sure whether this is representative for the gut responses.

A trend of negative association between *Firmicutes* and concentration of IFN-γ to PHA was seen in helminth-negative subjects only. In subjects with helminth, this association was positive, although this difference fell short of statistical significance. Parallel to this trend, the bacterial diversity was positively associated with IFN-γ responses to PHA in subjects who did not carry helminths, and in helminth-positive subjects, this association was dampened. Since a similar opposite trend was observed in the relationship between *Firmicutes* compared to bacterial diversity on IL-5 responses to PHA, we may conclude that not the proportion of *Firmicutes*, but the total bacterial diversity drove this association. *Firmicutes* was the most abundant phyla in this population and the increasing proportion of *Firmicutes* will obviously reduce diversity. This indicates that analyzing single bacterial phyla without considering the remaining phyla may lead to biased results as microbiome data is compositional and thus correlated between phyla.

Although deworming removed most helminths, treatment did not significantly alter the effects of bacterial proportions on cytokine responses. Regarding the *Bacteroidetes* effect on LPS to IL-10, we did observe a lower effect in the albendazole

group compared to placebo. Although not significant, this might point towards the idea that anthelminthic treatment could restore the possibly detrimental- interaction of bacteria with immune responses. Surprisingly, we found differences in immune modulation by *Actinobacteria* in the pre- versus post-treatment group. Although there was a significant association of time (in subjects receiving a placebo), these associations were not significantly different in the albendazole group. The effect of time could be explained by other factors such as diet and possibly improved hygiene, resulting from increased awareness during the presence of our medical team in the study area. In the analysis of treatment's effect on the association between bacterial proportion and diversity, there was a significant difference between the association of *Firmicutes* on the IL-5 response to PHA in the albendazole group compared to the placebo group. In subjects receiving albendazole, *Firmicutes* proportions were positively associated with IL-5 levels, while we observed a negative (non-significant) effect in helminth-negative individuals over time. This result seems contradictory, but might be related to the fact that small numbers were analyzed and not everyone in the albendazole group lost their helminth infection. The analysis of subjects who were infected at baseline and cleared their infection would possibly reveal more clearly how the relationship between bacterial communities and immunity are affected by treatment. This analysis lacks statistical power in our study as the sample size was small (n = 12 out of 17 subjects who were successfully dewormed). Future research which involves larger sample sizes needs to be conducted. With larger sample size, the current statistical model can be extended to account for different infection status at different time points as well as different randomization arms. Thereby we would gain more insight in the responses within individuals, and how these are affected by worm infection and deworming. Another relevant thought in this and similar research settings is that although albendazole removes helminths effectively, the immunomodulatory effects of helminths on cytokine responses are long-lasting and cannot be easily corrected by short-term treatment. It was previously reported by Endara et al. that the length of periodic treatment is important for altering immune responses (Endara et al., 2010), i.e. that studies with a longer period of treatment (up to 30 months) are more likely to show effects of deworming.

As significant associations between bacterial communities and cytokine responses were only observed when subjects were helminth-negative, clearly other factors than helminth and treatment are also involved in the alteration of the microbiome community and their interaction with the immune system. For example, our study data lack information on diet. Dietary intake was clearly shown to affect bacterial communities in the gastrointestinal tract (Wu et al., 2011). This might also be related to changes in social economic status leading towards a more high-fat diet when moving from rural to urbanized areas. Recent articles reported inconsistencies with regard to the direction of *Bacteroidetes* to *Firmicutes* ratio in rural to urban comparisons of microbiome profiles from different geographical areas. Studies comparing children from Bangladesh to USA children showed the direction of increasing Bacteroidetes: *Firmicutes* in the USA, as observed in our data (Lin et al., 2013), while studies in elderly

Korean and children in Burkina Faso showed opposing results, i.e. decreasing *Bacteroidetes*: *Firmicutes* ratios from rural to urban (Park et al., 2015; Filippo et al., 2017). This could be caused by different genera under *Bacteroidetes* or *Firmicutes* phyla which might be affected by certain type of diet. Therefore, it will be beneficial for future studies to also include dietary factors from the study participants.

A further limitation is related to the statistical tools available in analyzing this relationship. Here, we characterized the association of three single bacterial proportions on cytokine response in the helminth-positive and -negative group. Using this approach, we first ignore the effect of compositional structure in the microbiome data, namely when computing the *p*-value we assumed that these bacterial categories are independent while they are correlated. Secondly, the current statistical model ignores the fact that microbiome is a variable measured with errors at a different scale than the cytokine responses (Teixeira-Pinto et al., 2009). In addition, we might as well ignore the possible unobserved confounders. It is therefore important for future studies in this field to develop a statistical method to characterize the effects of helminth infection on both outcomes simultaneously by accounting these unobserved errors with a joint model.

To conclude, our findings support the hypothesis for the role of helminths in modulating the immune response, which might be related to bacterial proportion and diversity. Deworming did not show a particular effect on the observed associations. It is therefore important to repeat such studies with a larger sample size as well as using more advanced statistical models to further analyze this relationship by considering the complex structure of microbiome data and other possible confounders.

DATA AVAILABILITY STATEMENT

The microbiome datasets generated during the current study are available at the NCBI's Sequence Read Archive (SRA) accession numbers: SAMN07688522 to SAMN07688545. The ~4 million 16S assembled sequences from Indonesia samples are also available for download from Nematode.net (Nematode.net/Microbiome.html). The datasets supporting the conclusions of this article are available in the following Github address: https://github.com/IvonneMartin/Helminths_GutMicrobes_Cytokine.

ETHICS STATEMENT

This study was nested within the ImmunoSPIN study, a double-blind placebo-controlled trial conducted in Flores Island, Indonesia (Wiria et al., 2010). The ImmunoSPIN study has been approved by the Ethical Committee of Faculty of Medicine, Universitas Indonesia, ref:194/PT02.FK/Etik/2006 and has been filed by ethics committee of the Leiden University Medical Center. The clinical trial was registered with number: ISRCTN83830814. The subjects gave their informed consent either by written signature or thumb print. Parental consent was obtained for children below 15 years old.

AUTHOR CONTRIBUTIONS

IM, JH-D, and LW performed the analyses and wrote the manuscript. MK, AW, FH, YD, ES, and LW performed research on cytokine data. BR performed the microbiome data processing. TS, MM, and MY designed the study. All authors read, edited and approved the final manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.01028/full#supplementary-material

REFERENCES

A framework for human microbiome research. Nature. 2012;486:215–221. doi: 10.1038/nature11209

Abrams, G. D., Bauer, H., and Sprinz, H. (1963). Influence of the normal flora on mucosal morphology and cellular renewal in the ileum. A comparison of germ-free and conventional mice. *Lab. Invest.* 12, 355–364.

Bach, J.-F. (2002). The effect of infections on susceptibility to autoimmune and allergic diseases. *N. Engl. J. Med.* 347, 911–920. doi: 10.1056/NEJMra020100

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Soft.* 67 (1), 1–48. doi: 10.18637/jss.v067.i01

Belkaid, Y., and Hand, T. W. (2014). Role of the microbiota in immunity and inflammation. *Cell.* 157, 121–141. doi: 10.1016/j.cell.2014.03.011

Cooper, P., Walker, A. W., Reyes, J., Chico, M., Salter, S. J., Vaca, M., et al. (2013). Patent human infections with the whipworm, Trichuris trichiura, are not associated with alterations in the faecal microbiota. *PLoS ONE.* 8, e76573. doi: 10.1371/journal.pone.0076573

Endara, P., Vaca, M., Chico, M. E., Erazo, S., Oviedo, G., Quinzo, I., et al. (2010). Long-term periodic anthelmintic treatments are associated with increased allergen skin reactivity. *Clin. Exp. Allergy* 40, 1669–1677. doi: 10.1111/j.1365-2222.2010.03559.x

Filippo, C de, Di Paola, M., Ramazzotti, M., Albanese, D., Pieraccini, G., Banci, E., et al. (2017). Diet, Environments, and Gut Microbiota. A Preliminary Investigation in Children Living in Rural and Urban Burkina Faso and Italy. *Front. Microbiol.* 8, 1979. doi: 10.3389/fmicb.2017.01979

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *J. Stat. Soft.* 82 (13), 1–26. doi: 10.18637/jss.v082.i13

Lee, S. C., Tang, M. S., Lim, Y. A. L., Choy, S. H., Kurtz, Z. D., Cox, L. M., et al. (2014). Helminth colonization is associated with increased diversity of the gut microbiota. *PLoS Negl. Trop. Dis.* 8, e2880. doi: 10.1371/journal.pntd.0002880

Lin, A., Bik, E. M., Costello, E. K., Dethlefsen, L., Haque, R., Relman, D. A., et al. (2013). Distinct distal gut microbiome diversity and composition in healthy children from Bangladesh and the United States. *PLoS ONE.* 8, e53838. doi: 10.1371/journal.pone.0053838

Martin, I., Djuardi, Y., Sartono, E., Rosa, B. A., Supali, T., Mitreva, M., et al. (2018). Dynamic changes in human-gut microbiome in relation to a placebo-controlled anthelminthic trial in Indonesia. *PLoS Negl. Trop. Dis.* 12, e0006620. doi: 10.1371/journal.pntd.0006620

Mazmanian, S. K., Round, J. L., and Kasper, D. L. (2008). A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature.* 453, 620–625. doi: 10.1038/nature07008

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2017). *R-package: vegan: Community Ecology Package.* Vienna, Austria: R Foundation for Statistical Computing.

Park, S.-H., Kim, K.-A., Ahn, Y.-T., Jeong, J.-J., Huh, C.-S., and Kim, D.-H. (2015). Comparative analysis of gut microbiota in elderly people of urbanized towns and longevity villages. *BMC Microbiol.* 15, 49. doi: 10.1186/s12866-015-0386-8

Ramanan, D., Bowcutt, R., Lee, S. C., Tang, M. S., Kurtz, Z. D., Ding, Y., et al. (2016). Helminth infection promotes colonization resistance via type 2 immunity. *Sci.* 352, 608–612. doi: 10.1126/science.aaf3229

R Core Team. (2017). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing.

Reynolds, L. A., Finlay, B. B., and Maizels, R. M. (2015). Cohabitation in the Intestine: Interactions among Helminth Parasites, Bacterial Microbiota, and Host Immunity. *J. Immunol.* 195, 4059–4066. doi: 10.4049/jimmunol.1501432

Rook, G. A. W. (2009). Review series on helminths, immune modulation and the hygiene hypothesis: the broader implications of the hygiene hypothesis. *Immunology* 126, 3–11. doi: 10.1111/j.1365-2567.2008.03007.x

Rosa, B. A., Supali, T., Gankpala, L., Djuardi, Y., Sartono, E., Zhou, Y., et al. (2018). Differential human gut microbiome assemblages during soil-transmitted helminth infections in Indonesia and Liberia. *Microbiome* 6, 33. doi: 10.1186/s40168-018-0416-5

Round, J. L., and Mazmanian, S. K. (2009). The gut microbiota shapes intestinal immune responses during health and disease. *Nat. Rev. Immunol.* 9, 313–323. doi: 10.1038/nri2515

Teixeira-Pinto, A., Siddique, J., Gibbons, R., and Normand, S.-L. (2009). Statistical approaches to modeling multiple outcomes in psychiatric studies. *Psychiatr. Ann.* 39, 729–735. doi: 10.3928/00485713-20090625-08

Wammes, L. J., Mpairwe, H., Elliott, A. M., and Yazdanbakhsh, M. (2014). Helminth therapy or elimination: epidemiological, immunological, and clinical considerations. *Lancet Infect. Dis.* 14, 1150–1162. doi: 10.1016/S1473-3099(14)70771-6

Wammes, L. J., Hamid, F., Wiria, A. E., May, L., Kaisar, M. M. M., Prasetyani-Gieseler, M. A., et al. (2016). Community deworming alleviates geohelminth-induced immune hyporesponsiveness. *Proc. Natl. Acad. Sci. U. S. A.* 113, 12526–12531. doi: 10.1073/pnas.1604570113

Wexler, H. M. (2007). Bacteroides: the good, the bad, and the nitty-gritty. *Clin. Microbiol. Rev.* 20, 593–621. doi: 10.1128/CMR.00008-07

Wilson, M. S., Taylor, M. D., Balic, A., Finney, C. A. M., Lamb, J. R., and Maizels, R. M. (2005). Suppression of allergic airway inflammation by helminth-induced regulatory T cells. *J. Exp. Med.* 202, 1199–1212. doi: 10.1084/jem.20042572

Wiria, A. E., Prasetyani, M. A., Hamid, F., Wammes, L. J., Lell, B., Ariawan, I., et al. (2010). 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.* 10, 77. doi: 10.1186/1471-2334-10-77

Wu, G. D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.-Y., Keilbaugh, S. A., et al. (2011). Linking long-term dietary patterns with gut microbial enterotypes. *Science* 334, 105–108. doi: 10.1126/science.1208344

Xie, Y. *knitr: A General-Purpose Package for Dynamic Report Generation in R*, 2017. Vienna, Austria: R Foundation for Statistical Computing

Yazdanbakhsh, M., Kremsner, P. G., and van Ree, R. (2002). Allergy, parasites, and the hygiene hypothesis. *Science* 296, 490–494. doi: 10.1126/science.296.5567.490

Zaiss, M. M., Rapin, A., Lebon, L., Dubey, L. K., Mosconi, I., Sarter, K., et al. (2015). The intestinal microbiota contributes to the ability of helminths to modulate allergic inflammation. *Immunity* 43, 998–1010. doi: 10.1016/j.immuni.2015.09.012

Zhou, Y., and Zhi, F. (2016). Lower level of bacteroides in the gut microbiota is associated with inflammatory bowel disease: a meta-ssanalysis. *Biomed. Res. Int.* 2016, 5828959. doi: 10.1155/2016/5828959

Genomic Epidemiology in Filarial Nematodes: Transforming the Basis for Elimination Program Decisions

Shannon M. Hedtke[1]*, Annette C. Kuesel[2], Katie E. Crawford[1], Patricia M. Graves[3], Michel Boussinesq[4], Colleen L. Lau[5], Daniel A. Boakye[6] and Warwick N. Grant[1]

[1] Department of Physiology, Anatomy and Microbiology, La Trobe University, Bundoora, VIC, Australia, [2] Unicef/UNDP/World Bank/World Health Organization Special Programme for Research and Training in Tropical Diseases (TDR), World Health Organization, Geneva, Switzerland, [3] College of Public Health, Medical and Veterinary Sciences, James Cook University, Cairns, QLD, Australia, [4] Unité Mixte Internationale 233 "TransVIHMI", Institut de Recherche pour le Développement (IRD), INSERM U1175, University of Montpellier, Montpellier, France, [5] Department of Global Health, Research School of Population Health, Australian National University, Acton, ACT, Australia, [6] Parasitology Department, Noguchi Memorial Institute for Medical Research, Accra, Ghana

Onchocerciasis and lymphatic filariasis are targeted for elimination, primarily using mass drug administration at the country and community levels. Elimination of transmission is the onchocerciasis target and global elimination as a public health problem is the end point for lymphatic filariasis. Where program duration, treatment coverage, and compliance are sufficiently high, elimination is achievable for both parasites within defined geographic areas. However, transmission has re-emerged after apparent elimination in some areas, and in others has continued despite years of mass drug treatment. A critical question is whether this re-emergence and/or persistence of transmission is due to persistence of local parasites—i.e., the result of insufficient duration or drug coverage, poor parasite response to the drugs, or inadequate methods of assessment and/or criteria for determining when to stop treatment—or due to re-introduction of parasites *via* human or vector movement from another endemic area. We review recent genetics-based research exploring these questions in *Onchocerca volvulus*, the filarial nematode that causes onchocerciasis, and *Wuchereria bancrofti*, the major pathogen for lymphatic filariasis. We focus in particular on the combination of genomic epidemiology and genome-wide associations to delineate transmission zones and distinguish between local and introduced parasites as the source of resurgence or continuing transmission, and to identify genetic markers associated with parasite response to chemotherapy. Our ultimate goal is to assist elimination efforts by developing easy-to-use tools that incorporate genetic information about transmission and drug response for more effective mass drug distribution, surveillance strategies, and decisions on when to stop interventions to improve sustainability of elimination.

Keywords: population genomics, onchocerciasis, lymphatic filariasis, transmission, parasite elimination, drug resistance, epidemiology

INTRODUCTION

Onchocerciasis and lymphatic filariasis (LF) are targeted for elimination, primarily using community-wide mass drug administration (MDA). In April 2019, the World Health Organization (WHO) initiated a global consultation on the "Roadmap for Neglected Tropical Diseases," including current targets for global elimination of LF as a public health problem by 2020 (World Health Organozation, 2012a) and elimination of transmission of the parasite that causes onchocerciasis in 80% of affected sub-Saharan countries by 2025 (African Programme for Onchocerciasis Control, 2012). Barriers to elimination include operational challenges for adequate drug distribution or non-compliance leading to low treatment coverage (Churcher et al., 2009; Turner et al., 2013; African Programme for Onchocerciasis Control, 2015), movement of humans or vectors among communities (De Sole et al., 1991; African Programme for Onchocerciasis Control and World Health Organization, 2010; Ramaiah, 2013; Xu et al., 2018b; *Delineating Transmission Zones for Sustainable Onchocerciasis and LF Elimination*), and challenges in determining when the transmission cycle has been interrupted and treatment can be stopped (Stolk et al., 2015; Lont et al., 2017; Walker et al., 2017). Elimination of these diseases has been formally verified or validated by the WHO in some geographic regions, and is under assessment in others (*Parasite Transmission, Morbidity, and Control/Elimination Strategies*). However, there are areas where transmission has re-emerged after apparent (though not WHO-certified) elimination, and others where transmission has continued despite decades of MDA. A critical question is whether "local" parasites continue to persist because of insufficient MDA coverage and/or duration relative to the pre-intervention endemicity, poor parasite drug response, or inadequate methods of assessment and/or criteria for decisions to stop treatment, or whether "immigrant" parasites have been introduced *via* human or vector movement from another endemic area. Population genetic analyses of parasite DNA sequences may provide data to answer this question. With the increasing technical feasibility of gathering population-level genomic data, it becomes possible to develop tools allowing routine acquisition of such data by elimination programs to inform decisions on MDA strategies, delineate areas to be included in evaluations, and determine whether treatment should be continued or stopped.
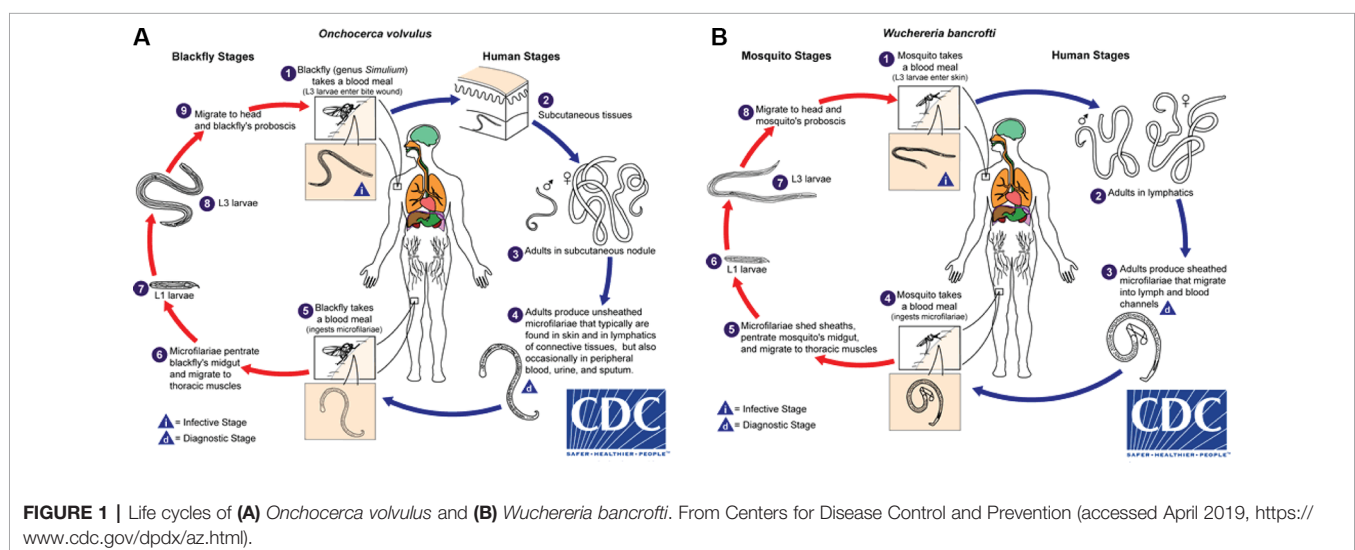
Parasite Transmission, Morbidity, and Control/Elimination Strategies
Onchocerciasis

Onchocerca volvulus is transmitted by blackflies of the genus *Simulium*. Female blackflies taking a blood meal from an infected person ingest *O. volvulus* microfilariae, which develop into infective L3 larvae that are transmitted to a different host during subsequent blood meals (**Figure 1A**). The L3 larvae develop *via* L4 into adult worms, macrofilariae, which live for around 12–14 years in nodules under the skin and deep in the body, producing millions of microfilariae. The microfilariae migrate through the skin, eyes, and other organs, and live for up to 2 years. The immunological reactions upon death of microfilariae are the major cause of morbidity: primarily itching, skin depigmentation and lesions, and visual impairment that can progress to blindness (Remme et al., 2017). Increasing evidence suggests that high *O. volvulus* infection can induce epilepsy (Chesnais et al., 2018).

The pattern of onchocerciasis endemicity is determined by ecological conditions favorable for *Simulium* breeding sites, whose location and productivity (i.e., number of blackflies hatching) can vary seasonally and from year to year. *Simulium damnosum sensu lato* (*s.l.*), the main vector in Africa, breeds in fast-flowing rivers and streams with the high level of oxygenation needed for larval development. The flies typically have an active flight range of around 15 km, but can migrate seasonally over hundreds of kilometers if assisted by prevailing winds, and thus transmit the parasite over large geographic ranges (Garms et al., 1979; Le Berre et al., 1979; Baker et al., 1990). Infection rates are highest among those living and/or working close to consistently highly productive breeding sites and decline with increasing

FIGURE 1 | Life cycles of **(A)** *Onchocerca volvulus* and **(B)** *Wuchereria bancrofti*. From Centers for Disease Control and Prevention (accessed April 2019, https://www.cdc.gov/dpdx/az.html).

distance between where blackflies breed and where people live and work.

The morbidity and socio-economic impact of onchocerciasis motivated three control programs. The Onchocerciasis Control Programme in West Africa (OCP, 1974–2002) targeted elimination as a public health problem through weekly aerial larviciding of breeding sites along ~50,000 km of rivers for the life span of macrofilariae, later complemented in some areas by mass drug administration of ivermectin (MDAi). The OCP originally covered ~654,000 km^2, which was expanded to 1,300,000 km^2 because *Simulium* from outside invaded the original area (Le Berre et al., 1979; Baker et al., 1990; Remme, 2004; Boatin, 2008).
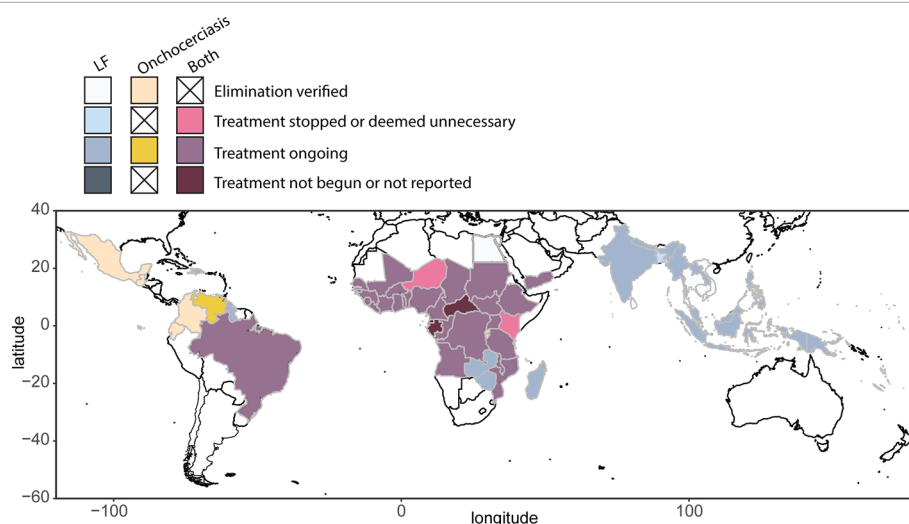
The 1987 decision of Merck & Co., Inc. (Kenilworth, NJ, USA) to donate ivermectin (Mectizan®) for onchocerciasis control, and subsequent studies showing the safety of MDAi, allowed initiation of MDAi-based control programs. The Onchocerciasis Elimination Program for the Americas (OEPA) was initiated in 1993 to eliminate transmission in 13 generally small foci across six countries with a total at-risk population of 0.56 million. Building on early efforts to control onchocerciasis through nodulectomy campaigns and annual ivermectin treatments, the OEPA implemented health system–directed biannual MDAi (and even quarterly MDAi in four areas in Brazil, Mexico, and Venezuela). Elimination was certified by WHO for all but the large Amazonian area across the border between Brazil and Venezuela, where difficult terrain and lack of roads make it challenging to ensure ivermectin distribution to ~30,500 Yanomami across ~540 migratory communities (Sauerbrey et al., 2018; World Health Organization, 2018; **Figure 2**).
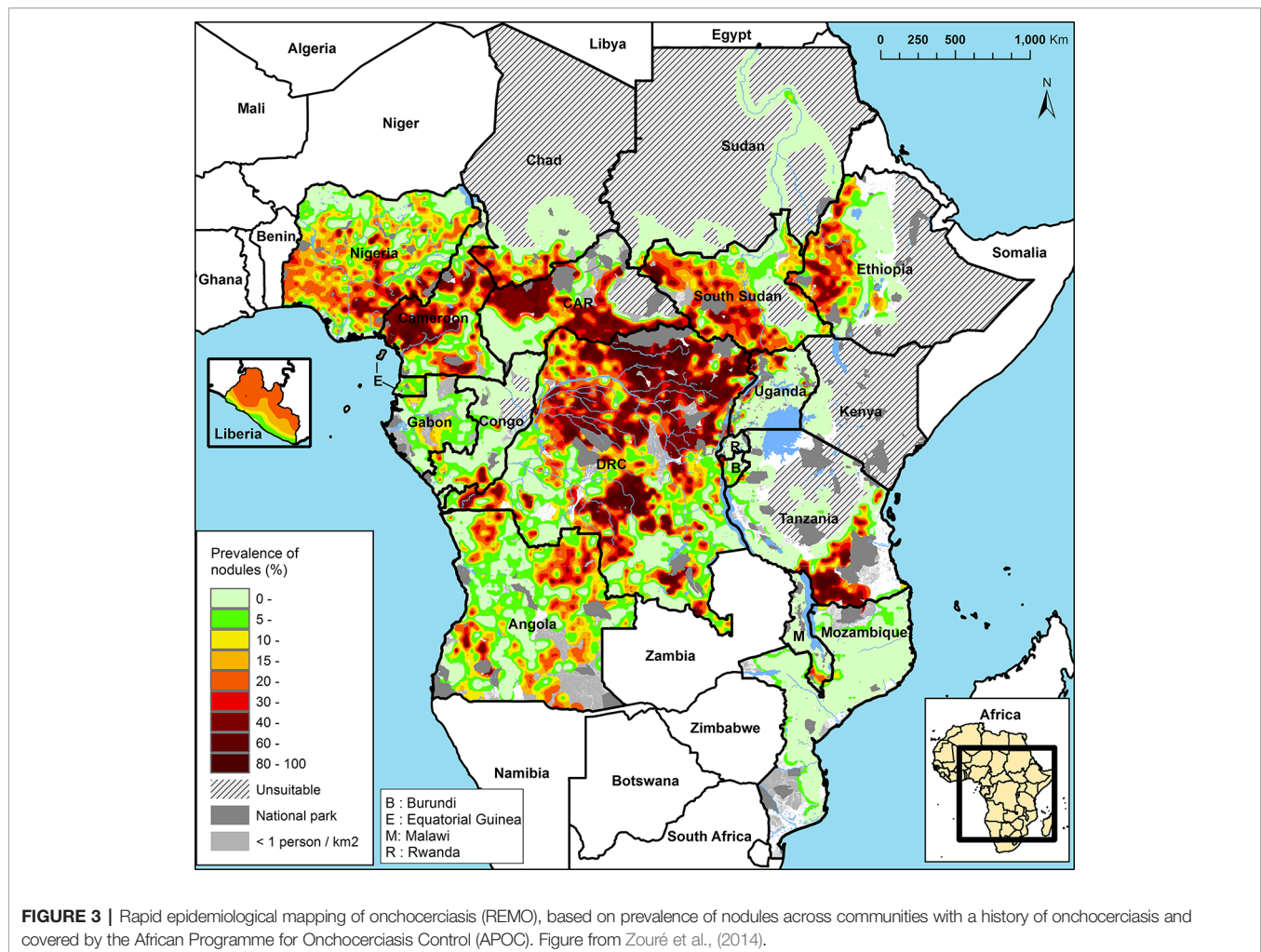
The APOC (1995–2015) initially targeted control of onchocerciasis as a public health problem in Central and East Africa and Liberia with sustainable, community-directed ivermectin treatment (CDTI). Across the 19 APOC countries, ~86 million people were estimated to require CDTI because they lived in meso- and hyperendemic areas, i.e., where 20–40% and >40% of adults, respectively, have subcutaneous nodules (Noma et al., 2014; **Figure 3**). These nodule prevalences correspond to prevalences of skin microfilariae in the general population of around 40–60% for mesoendemic and >60% for hyperendemic areas associated with an increased risk of onchocercal blindness (Prost et al., 1979; UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases, 1992; Noma et al., 2002; Seketeli, 2002; Seketeli et al., 2002; Fobi et al., 2015). The sizes of these areas range from relatively small to a vast contiguous endemic area of ~2 million km^2 across seven countries (Noma et al., 2014; World Health Organization, 2018). Research and subsequent epidemiological evaluations of infection rates in areas with long-term MDAi (Diawara et al., 2009; Traore et al., 2012; Tekle et al., 2016) suggested that MDAi alone could achieve elimination of transmission in Africa. The APOC objectives were consequently expanded to onchocerciasis elimination in at least 80% of African countries by 2025 (African Programme Onchocerciasis Control, 2012). In 2015, APOC was closed. The Expanded Special Project for Elimination of Neglected Tropical Diseases (ESPEN) now provides countries with some of the support previously provided by APOC.

MDAi in meso- or hyperendemic areas may have reduced infection prevalence in neighboring hypoendemic areas. Previously unmapped hypoendemic areas that might impact elimination goals (Rebollo et al., 2018) by acting as a potential source of renewed transmission need to be efficiently determined and cost-effective intervention strategies implemented in these areas (and those previously identified as hypoendemic and not included in CDTI). Including the population in hypoendemic areas mapped by 2017–2018, WHO estimated that 204 million people required MDAi to achieve elimination (World Health

FIGURE 2 | Global distribution and treatment of lymphatic filariasis (2018) and onchocerciasis (2017), based on WHO Global Health Observatory data repository (accessed April 2019, https://www.who.int/gho/en/).

FIGURE 3 | Rapid epidemiological mapping of onchocerciasis (REMO), based on prevalence of nodules across communities with a history of onchocerciasis and covered by the African Programme for Onchocerciasis Control (APOC). Figure from Zouré et al., (2014).

Organization, 2018). In Central Africa where another filarial parasite, *Loa loa*, is endemic, areas previously identified as requiring interventions based on parasites observed in the skin may require remapping. Recent data suggest that *L. loa* microfilariae, once assumed to be present only in blood, may also be present in the skin and mistaken for *O. volvulus* (Nana-Djeunga et al., 2019).

The OCP experience (Remme et al., 1995) informed the 2001 WHO procedures for entomological assessments and criteria for certification of onchocerciasis elimination (World Health Organization, 2001). Those procedures stipulated testing at least 10,000 flies from each endemic community selected for evaluation by the WHO International Certification Team, maximizing the chances of detecting infected flies by collecting during peak transmission season. In 2010, APOC outlined a "Conceptual Framework of Onchocerciasis Elimination with Ivermectin Treatment" which built on the OCP and APOC experience, took into account differences between control *via* vector elimination and *via* ivermectin treatment, and emphasized that much is still to be learned (African Programme for Onchocerciasis Control/World Health Organization, 2010; Dadzie et al., 2018). This framework

defined elimination as "reduction of infection and transmission to the extent that interventions can be stopped but post-intervention surveillance is still necessary" and a transmission zone as a "geographical area where transmission of *O. volvulus* occurs by locally breeding vectors and which can be regarded as a natural ecological and epidemiological unit for interventions." It identified delineating transmission zones and examining risks of parasite reintroduction through human or vector migration as critical to decisions to stop MDA. Acknowledging that "in practice it is difficult to determine with a fair degree of certainty that vectors in a given area are exclusively locally breeding," the framework recommended an operational definition of a transmission zone as a "river basin, or a major section of a river basin, where onchocerciasis is endemic and where the river is the core of the endemic area, with communities with the highest prevalence of infection generally located close to the river and infection levels falling with increasing distance from the river until they become negligible or reach a neighboring transmission zone." In 2016, the WHO issued a new guideline for stopping MDA and verifying elimination of human onchocerciasis (World Health Organization, 2016). This guideline defines interruption of transmission of *O. volvulus* as

"the permanent reduction of transmission in a defined geographical area after all the adult worms (and microfilariae) in the human population in that area have died, been exterminated by some other intervention, or become sterile and infertile." While this guideline does not specify criteria for a "defined geographical area," it defines a transmission zone in the same way as the APOC framework. The 2016 Guideline Development Group recommended that decisions about interruption of parasite transmission, stopping MDA, and confirmation of transmission interruption at the end of the post-treatment surveillance period should be based on a prevalence of infective parous flies of <0.1%. Statistical confidence for this elimination threshold requires that a minimum of 6,000 flies collected from "a transmission zone" should be tested. The guideline does not address delineation of a transmission zone and hence the geographical range over which those flies are to be collected. To date, no validated methodology for transmission zone delineation is available. We review evidence here that population genetic measures of the parasite (and its vectors) capture the history of transmission and hence can be used as a proxy for defining transmission zone boundaries.

Lymphatic Filariasis

LF is a vector-borne disease caused by three species of parasites: *Wuchereria bancrofti*, *Brugia malayi,* and *Brugia timori*. The majority of infections are due to *W. bancrofti*, which is found throughout the tropics (**Figure 2**). Chronic infection with these parasites can disrupt lymphatic function and lead to long-term disability due to lymphedema (most commonly swollen lower limbs) and scrotal hydrocele (Taylor et al., 2010; World Health Organzation, 2011). Globally, ~68 million are infected with LF, with 36 million microfilaremic persons, 19 million hydrocele cases, and 17 million lymphedema cases (Ramaiah and Ottesen, 2014). In 1997, following a World Health Assembly resolution, the WHO initiated elimination of LF as a public health problem, using a two-pillar program of annual MDA to reduce microfilaremia to levels that would not sustain continued transmission, and management of morbidity and disability by providing a minimum package of care (Ichimori et al., 2014).

Infective larval LF are transmitted to humans by mosquitoes and migrate to the lymphatic system where they develop into adult worms that live for ~4 to 6 years (**Figure 1B**). After mating, female adult worms produce thousands of microfilariae throughout their life, which circulate in the peripheral blood and are available to infect mosquitoes during blood meals to continue the transmission cycle. LF is transmitted by a wide range of mosquito genera including *Anopheles*, *Culex*, *Mansonia*, and *Aedes* (World Health Organization, 2013). In a striking example of the power of adaptive evolutionary change, the daily periodicity in density of microfilariae in peripheral blood is tuned to the biting times of local vectors.

Human infections are diagnosed by detecting microfilariae on blood films collected at the appropriate time of day, or by detecting circulating filarial antibodies or antigens (Ag). Detecting microfilariae on blood films requires technical skill

and equipment and collecting blood at night (where vectors are nocturnal) is challenging, so serology is usually used rather than microscopy. The current standard test for programmatic surveillance activities in *W. bancrofti* areas is a rapid diagnostic test for circulating filarial antigens (World Health Organization, 2017b). Antibody tests are not usually used because antibodies can persist for years after the death of adult worms. Both overall and age-specific prevalence of LF infection (usually not peaking until adulthood) is determined by transmission intensity, which in turn depends on vector biting rates and the availability of susceptible humans. The threshold for starting MDA programs is microfilariae or antigen prevalence of ≥1% within an implementation unit (IU), which is defined prior to mapping (World Health Organization, 2011; World Health Organization, 2017b). The size of the IU was left to the judgment of the endemic countries; some chose to map by village or district, while others classified the entire country or regions. In areas in Africa co-endemic for loiasis (*L. loa* filariasis), the use of antigen-based test strips for mapping has probably overestimated LF prevalence, because of false positives from people with high intensity infection with *L. loa* (Wanji et al., 2015; Pion et al., 2016). Thus, mapping should be revisited in loiasis co-endemic areas using approaches that increase species specificity.

MDA is conducted annually with a single-dose combination of albendazole with diethylcarbamazine (DEC) or, in onchocerciasis co-endemic countries, with ivermectin. Recently, the three drugs have been combined in some countries (King et al., 2018), including those where elimination targets have not been reached despite completing the required rounds of two-drug MDA (World Health Organization, 2017a). In loiasis co-endemic areas in Central Africa not meso- or hyperendemic for onchocerciasis, biannual albendazole treatment in combination with vector management is recommended, since DEC and ivermectin can cause serious adverse reactions in individuals with high *L. loa* microfilaremia (Boussinesq et al., 2003; World Health Organization, 2012b).

The WHO recommendation is for MDA to be conducted until large, well-designed transmission assessment surveys (TASs) of 6- to 7-year-old children in schools or communities determine that the estimated antigen prevalence in an evaluation unit has dropped below a critical cutoff value, when transmission is not considered sustainable (World Health Organization, 2011; World Health Organization, 2017b). Critical cutoff values depend on the filarial and vector species, and are calculated such that the likelihood of an evaluation unit passing is at least 75% if true antigen prevalence is 0.5%, and no more than 5% if the true antigen prevalence is 1% (World Health Organization, 2011; World Health Organization, 2017b). For example, in regions where *W. bancrofti* is endemic and transmission is dominated by *Aedes* spp. mosquitoes, the threshold is based on an upper 95% confidence limit for antigen prevalence of 1%. The TAS evaluation unit can be the same as the IU, it can be larger and include several IUs, or it can be smaller (by splitting an IU), depending on available information on likely residual LF endemicity. Once a country has passed two post-MDA TASs at 2- to 3-year intervals in all evaluation units, it can apply for

validation of elimination of LF as a public health problem. Post-validation surveillance also uses TAS in 6- to 7-year olds (World Health Organization, 2017b). Morbidity management and rehabilitation must continue as needed.

Delineating Transmission Zones for Sustainable Onchocerciasis and LF Elimination

Onchocerciasis

The mosaic of onchocerciasis hyper-, meso-, and hypoendemicity in Africa implies an underlying geographic mosaic of parasite transmission zones (**Figure 3**; African Programme for Onchocerciasis Control/World Health Organization, 2010) that is the product of long-term, historical spatial density and migration patterns of the blackfly vector and of the human host (Blouin et al., 1995; Nadler, 1995; Jarne and Théron, 2001; Criscione and Blouin, 2004; Criscione et al., 2005; Prugnolle et al., 2005; Barrett et al., 2008). This pattern was not important for implementation of "project areas" for control of onchocerciasis as a public health problem (i.e., morbidity reduction and prevention), which were delineated based on the administrative borders of health system units in which meso- and hyperendemic areas were located. However, with the switch from control as a public health burden to elimination of transmission, explicit consideration should be given to the alignment of intervention areas and the areas included in the evaluations for decisions to stop treatment (subsequently referred to as "evaluation areas") with the true, long-term parasite transmission zones.

Although the success of MDAi has driven the move from control to elimination of onchocerciasis, there are some areas in Africa in which many years of MDAi have reduced infection prevalence less than expected (i.e., less than transmission models predict) or where resumption of transmission has been reported following cessation of interventions (Kamga et al., 2016; Tekle et al., 2016; Koala et al., 2017; Bakajika et al., 2018; Komlan et al., 2018; Koala et al., 2019). Transmission models predict that the duration of interventions required to interrupt transmission is dependent on vector abundance (which determines the percentage of the population infected before interventions begin), and the treatment frequency and percentage of the population taking ivermectin (African Programme for Onchocerciasis Control/World Health Organization, 2010; Stolk et al., 2015). Consequently, the apparent failure of MDAi to interrupt transmission in some treatment areas, and of post-treatment recrudescence in others, might be due to highly hyperendemic areas requiring longer-than-assumed duration of MDAi (and/or higher treatment coverage and/or more frequent treatment; Stolk et al., 2015) or higher-than-appropriate thresholds of vector infectivity for deciding to stop interventions. An alternative explanation that has yet to be modeled, or considered explicitly in any guideline, is that continuing transmission/recrudescence may be a result of continuing parasite invasion from neighboring areas; i.e., there may be discordance between an intervention or evaluation area and the parasite transmission zone within which the area is located. In order to achieve sustainable elimination, transmission must be interrupted in the whole transmission zone, which needs to be effectively delineated to ensure that the whole zone is included in both interventions and evaluations.

The risk of parasite invasion is illustrated by two reports from river basins in Burkina Faso where transmission was considered to have been permanently interrupted. In the Comoé basin, retrospective analysis of recrudescence suggested that long-range vector movement was responsible, although immigration of infected people from outside the basin was not excluded. Ongoing transmission was confirmed *via* entomological assessments at 4/4 capture points (Koala et al., 2017; Koala et al., 2019). The second Burkina Faso report investigated microfilariae-positive individuals in the Upper Mouhoun, Nakambé, and Nazinon river basins and found all but 1 of the 31 infected individuals recently resided in neighboring Côte d'Ivoire (Nikièma et al., 2018). The remaining person had been previously identified as infected in 2008 (Nikièma et al., 2018). Thus, migration between an area with and an area without ongoing transmission accounted for 30/31 of the identified infected people.

The two cases described above point to immigrant parasites as the potential source of renewed transmission and emphasize the importance of knowing whether or not new infections are due to parasites which "immigrated" *via* vectors or humans for determining areas to be included in appropriate treatment strategies. Similarly, the estimated risk of immigration *via* humans or vectors is important for elimination programs to make cost–risk based decisions: i.e., evaluate the cost of stopping MDAi based on the risk of resurgence due to immigration from areas that do not yet meet stopping criteria versus the cost of continuing MDAi until these other areas also meet stopping criteria. Finally, such information can provide lessons for evaluations required for decisions to stop interventions, including delineation of evaluation areas, sampling criteria, or thresholds for MDAi cessation, which are all currently determined independent of vector abundance.

Lymphatic Filariasis

In 2012, 73 countries were regarded as LF endemic (Ramaiah and Ottesen, 2014), with 55 implementing MDA. By October 2019, 15 countries had been validated as having eliminated LF as a public health problem: Vanuatu, Tonga, Niue, Cook Islands, Republic of the Marshall Islands, Palau, Wallis & Futuna, Kiribati, Cambodia, and Vietnam in the WHO Western Pacific Region (WPRO); Sri Lanka, Maldives, and Thailand in the WHO South East Asian Region (SEARO); Egypt in the WHO Eastern Mediterranean Region (EMRO); and Togo in the WHO African Region (AFRO) (World Health Organization, 2019) (**Figure 2**).

While 15 countries have achieved validation of elimination so far (and some are nearing this milestone), there are countries where progress toward elimination has stagnated or reversed, including several in the Pacific where LF is transmitted by efficient day-biting *Aedes* vectors and where there may also be supplementary night-biting vectors (Ichimori and Graves, 2017). These countries include American Samoa, Samoa, French Polynesia, and Fiji.

LF is relatively heterogeneous in its distribution within countries, and thus the role of movement between

intervention/evaluation units in maintaining or re-establishing transmission is a concern for many national LF elimination programs. Transmission zones in *W. bancrofti* are highly dependent on variation in vector density, survival, and competence for transmission and whether vector mosquitoes are diurnal, nocturnal, or sub-nocturnal. In areas with post-intervention recrudescence, there are sometimes infection "hot spots": a localized area where prevalence is significantly higher than surrounding areas (e.g., American Samoa: Lau et al., 2014; Lau et al., 2016; Lau et al., 2017; Sri Lanka: Rao et al., 2014; Rao et al., 2017; Rao et al., 2018). Potential explanations for recrudescence include expansion of infections from hot spots when mosquitoes take up and transmit parasites to geographically nearby communities or when infectious people visit other communities, slow but insidious increase in frequency and intensity of transmission from widespread infections at low prevalence that were not detected or were considered (incorrectly) to be under the threshold for stopping MDA, or migrants from other endemic regions continue to introduce new parasites, if the evaluation unit is inappropriately small.

The role of population movement between countries and regions within countries in maintaining or re-establishing transmission is a concern for many national elimination programs. Continued surveillance post-MDA has been recommended by researchers when communities are in close proximity to cross-border regions with ongoing transmission (Ramaiah, 2013; Dorkenoo et al., 2018). The possibility that highly mobile migrant workers from Myanmar introduced and maintained LF transmission in Thailand, after effective MDA throughout most of the latter country, resulted in additional rounds of MDA in border communities and sparked epidemiological research on the connectivity between communities (Bhumiratana et al., 2010; Satimai et al., 2011; Bhumiratana et al., 2013; Toothong et al., 2015; Dickson et al., 2017). The influx of thousands of refugees from Haiti, where LF is hyperendemic, to Brazil, where LF transmission has been reduced and even eliminated in some states, has prompted screening and treatment of recent immigrants to reduce risk of re-introduction (Nunes et al., 2016; da Silva et al., 2017; Zuchi et al., 2017). Even in the remote islands of Samoa and American Samoa, models have shown that movement of people between countries could have a significant impact on the transmission of infectious diseases, including LF (Xu et al., 2018b). The role of movement of infected persons in the establishment and persistence of hot spots, and thus the appropriate choice of the size for implementation and evaluation units (both in terms of geographic size and total population) for LF programmatic activities needs to be critically evaluated for their impact on elimination progress. In contrast, long distance vector dispersal is not typically considered a major risk for increasing the distribution or prevalence of the LF parasite (Hapairai et al., 2013; Verdonschot and Besse-Lototskaya, 2014, but see Huestis et al., 2019).

Crucial questions to be addressed for sustainable LF elimination are thus: (1) what drives development of hot spots, and their persistence after MDA, (2) what are effective methods for post-MDA surveillance, when very low infection prevalence becomes difficult to detect, (3) how to determine appropriate target thresholds for successful interruption of transmission in different settings (e.g., areas with different mosquito vectors), (4) what strategies should be used to delineate intervention/evaluation units, (5) what role does travel/migration (i.e., population movement and connectedness) play in the transmission of parasites between countries and between regions within countries, and (6) how do transmission dynamics among regions affect planning for future MDA? To stop interventions and enter the "post-MDA surveillance" phase in a specific geographic area requires reasonable certainty that the probability of importation of parasites is too low to re-initiate transmission, i.e., that stopping criteria are met across the whole transmission zone. In other words, stopping MDA for LF also requires delineation of transmission zone boundaries.

PARASITE POPULATION GENETICS AND TRANSMISSION ZONES

Filarial nematodes have nuclear genomes on the order of ~90–100 Mbp, mitochondrial genomes of ~13–14 Kbp, and most, but not all, species obligately carry a *Wolbachia* endosymbiont that has its own genome of ~1 Mbp (Keddie et al., 1998; Keddie et al., 1999; Unnasch and Williams, 2000; Foster et al., 2005; McNulty et al., 2012; Ramesh et al., 2012; Desjardin et al., 2013; Cotton et al., 2016; Small et al., 2016; Chung et al., 2017). Sequencing, assembling, and annotating all three genomes for many helminth species, including the filariae, is the objective of continuing efforts, and the availability and assembly quality of parasite genomes will likely continue to increase rapidly. Currently, the *O. volvulus* genome has the highest quality assembly, with complete assemblies of mitochondrial and *Wolbachia* genomes and nuclear genome sequence scaffolds that are approaching whole chromosomes. Genome resources for this and other parasites can be found at https://parasite.wormbase.org/index.html.

With a few notable exceptions (Choi et al., 2016; Small et al., 2016; Doyle et al., 2017; Small et al., 2019), most genomic research for filarial nematodes has focused on predictions of gene function. The aim of many of these "functional" predictions is to identify targets for new drug discovery, usually in the absence of methods for functional genomic validation of targets. Given the time for moving from identification of potential drug targets to regulatory registration and WHO guidelines for new drugs, it is unclear to what extent drugs emerging from such research are going to be available for use by LF and onchocerciasis elimination programs.

In contrast, relatively little attention has been paid to the potential for the explosion of genomic information to support elimination efforts with *currently used* drugs or new applications of drugs such as moxidectin, which achieved regulatory registration with the U.S. Food and Drug Administration in 2018 and is undergoing additional studies for inclusion in WHO guidelines (Opoku et al., 2018). Understanding the geographic distribution of variation in susceptibility to those drugs—whether, for example, drug-naive populations already contain alleles for reduced susceptibility—may allow us to act to preserve susceptibility to those drugs. Population genomics analyzes

whole-genome variation within and between parasite populations, identifying where that variation might impact drug efficacy (the evolution of drug resistance) or elucidating patterns of disease transmission (genetic epidemiology). We review applications of population genomics to informing when and where to cease treatment, on the quantifiable risk of recrudescence through migration of infected humans/vectors from other areas (*O. volvulus Population Genetic Structure* and *W. bancrofti Population Genetic Structure*), and on the long-term sustainability of MDA if poor response to drugs is heritable (*Parasite Population Genomics for Understanding Geographic and Temporal Variability in Drug Response*).

Conceptual Approach

Genetic structure is the result of movement of parasites within, but not between, transmission zones: parasites within a transmission zone interbreed and are genetically more similar to each other than to parasites in another transmission zone, with whom they do not interbreed. The likelihood that parasites are from the same transmission zone, and thus likelihood of parasite transmission between two locations (invasion), can therefore be inferred from the degree of genetic relatedness between them (e.g., Blouin et al., 1995; Nadler, 1995; Criscione and Blouin, 2004; Small et al., 2014). Statistically significant genetic differentiation at neutral loci (i.e., loci which do not affect a phenotype subject to selection pressure) between two parasite samples indicates that they originate from two separate transmission zones (**Figure 4**). In contrast, high genetic similarity among parasites indicates either historical or ongoing gene flow (i.e., interbreeding), and that these more closely related parasites are from a single transmission zone, regardless of geographical scale. This genetic population structure—or lack of it—is the result of long-term patterns of transmission. Measures of population genetic structure could thus help programs assess the risk of resurgence associated with alternative scenarios on a longer time scale than intrinsically shorter-term decisions about where and whom to treat. Population genetic measures have proven to be informative for malaria epidemiology and control, particularly for investigating sources of transmission or the origins and prevalence of drug resistant strains (Jennison et al., 2015; Lo et al., 2017; Aydemir et al., 2018; Fola et al., 2018; Waltmann et al., 2018).

In population genetic analyses, sequence data are collected from parasites from multiple geographic locations. One approach to analyzing these data is to compare and cluster sequences based on the distribution of genetic variation across a geographic area: members of the same parasite population share similar patterns of genetic sequence variation because they have been interbreeding (e.g., Pritchard et al., 2000; Jombart et al., 2010). An alternative approach is to explore the history of parasite movement by simulating the expected distribution of genetic variation under different parasite migration histories, and comparing test statistics derived from these simulated data to test statistics calculated using the observed genetic variation (e.g., Laval and Excoffier, 2004; Peng and Kimmel, 2005). Development of analytical methods suitable for quantitative assessment of transmission of parasites within and among geographic regio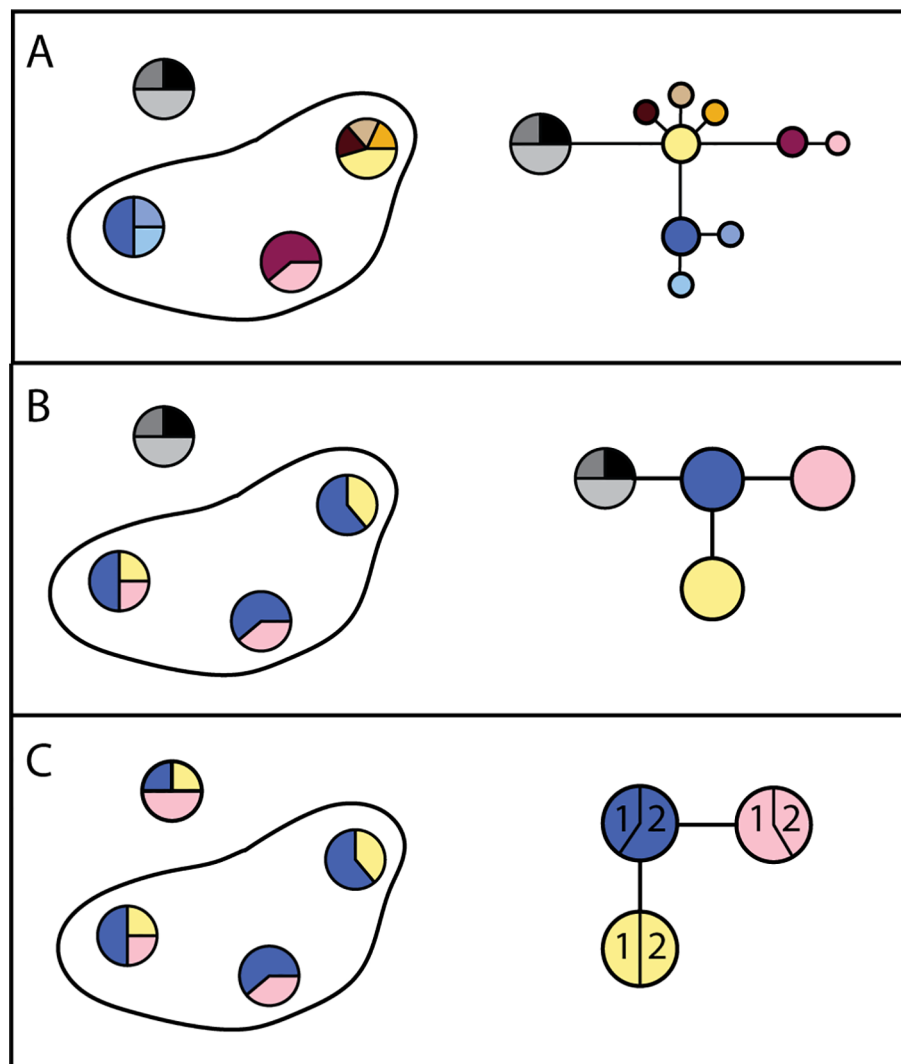ns is an active area of research, and more in-depth reviews of these methods can be found elsewhere (e.g., Schawbl et al., 2017; Wesolowski et al., 2018).

Prior to MDA, filarial parasite populations had a widespread, continuous distribution (e.g., onchocerciasis in Central/East Africa; **Figure 3**). Because MDA (or vector control) reduces or interrupts transmission, parasite population sizes fluctuate over time, and their geographic distribution becomes increasingly patchy. Parasite movement within a large geographic area in which parasite populations were reduced to zero only in some subareas can lead to transmission being re-established where it was previously interrupted. These re-established parasite populations and the parasite populations from which they derive form what in ecology and population genetics is called a "metapopulation" (Levins, 1969; Ariey et al., 2003; Churcher et al., 2008; Tadiri et al., 2018). Where there is a pattern of re-establishment post-intervention, parasites within a metapopulation are genetically similar, and the geographic range of this metapopulation defines the transmission zone.

Parasitic nematodes tend to have higher genetic diversity than vertebrates (e.g., Blouin et al., 1992; Grant and Whitington, 1994; Blouin et al., 1995; Hawdon et al., 2001; Prichard, 2001), possibly due to their complex life cycles, transmission epidemiology (Criscione et al., 2005), and large population size. Although the high genetic diversity in parasites may require a large sample size to ensure the data are representative, estimating population genetic structure does not require a complete, assembled genome. Researchers target genomic regions identified as sufficiently informative to reveal relatedness among individuals. For example, estimating genetic diversity and population structure using mitochondrial sequence data is well established in population genetics. The advantages of mitochondrial markers include that they are maternally inherited, have a higher mutation rate than the nuclear genome (10- to 17-fold; although this may be lower in species associated with *Wolbachia*; Cariou et al., 2017), have a high copy number relative to the nuclear genome in each cell (useful for DNA sequencing of samples with low concentration or degraded DNA), and do not generally recombine. Mitochondrial genomes upon which markers for population genetics can be built are also widely available across filarial nematodes, including *O. volvulus* (Keddie et al., 1998; Unnasch and Williams, 2000), *W. bancrofti* (Ramesh et al., 2012; Small et al., 2016), *B. malayi* (Williams et al., 2000), *L. loa* (Higazi et al., 2004), and the canine heartworm, *Dirofilaria immitis* (Hu et al., 2003).

There are advantages to conducting population genetic analyses that include nuclear markers: they are inherited from both parents, and the extent of recombination between markers can be informative about past demographic events such as reductions in population size or selection on heritable traits (both events cause markers to be linked: two or more DNA variants on a chromosome become associated more often than expected by chance). Nuclear variation has been estimated using partial sequences of multi-copy ribosomal genes or spacers, common repeat sequences, microsatellites, and fingerprinting using electrophoresis of randomly amplified DNA (RAPD). While these approaches have been useful due to their low cost compared to whole-genome sequencing, ease of experimental

FIGURE 4 | Alternate possibilities for distribution of genetic variation in parasites within a hypothetical country. **(A)** Haplotype frequencies (each unique sequence has a different color) from three locations within one country where nematodes were collected, and from a second country (haplotypes in grey), indicate all three locations are genetically distinct. The resulting haplotype network indicates sequence divergence between haplotypes. Taken together, these results would support the conclusion that there are three transmission zones in which parasites persist. **(B)** Haplotype frequencies in three locations where nematodes were collected indicate that they are genetically similar. The resulting haplotype network supports the conclusion that there is one transmission zone, and that recrudescence is local. **(C)** Haplotype frequencies in three locations indicate that they are genetically similar. The haplotype network indicates that recrudescence may have come from a migrant source, because haplotypes are shared between countries (1 and 2).

procedures, and straightforward comparisons between species (e.g., Blouin et al., 1995), capturing only a small part of genome variation underestimates diversity and can result in estimates of genetic variation with low information content (e.g., Keddie et al., 1999; Higazi et al., 2004). Analyses that utilize whole-genome sequencing data, rather than a few selected loci, can also identify the confounding effects of selection on estimates of population size, structure, and migration (see, e.g., Weigand and Leese, 2018). Increasing the information available per individual using whole-genome sequencing of either individuals or pools of individuals, and increasing the information available per population by increasing the total number of individuals sequenced, has improved the sensitivity

and power of analyses, and will improve our understanding of disease transmission.

O. volvulus Population Genetic Structure

O. volvulus population genetic structure is determined by the geographic range and movement of its hosts and vectors. The major blackfly vectors in Africa belong to a species complex (*S. damnosum s.l.*) divided into a number of subspecies based on cytotaxonomy (Boakye, 1993; Boakye and Meredith, 1993; Krueger and Hennings, 2006), although these subspecies are not all distinguishable using DNA sequence data (Tang et al., 1995; Tang et al., 1996). Variation in vector species distribution could influence *O. volvulus* population structure, because
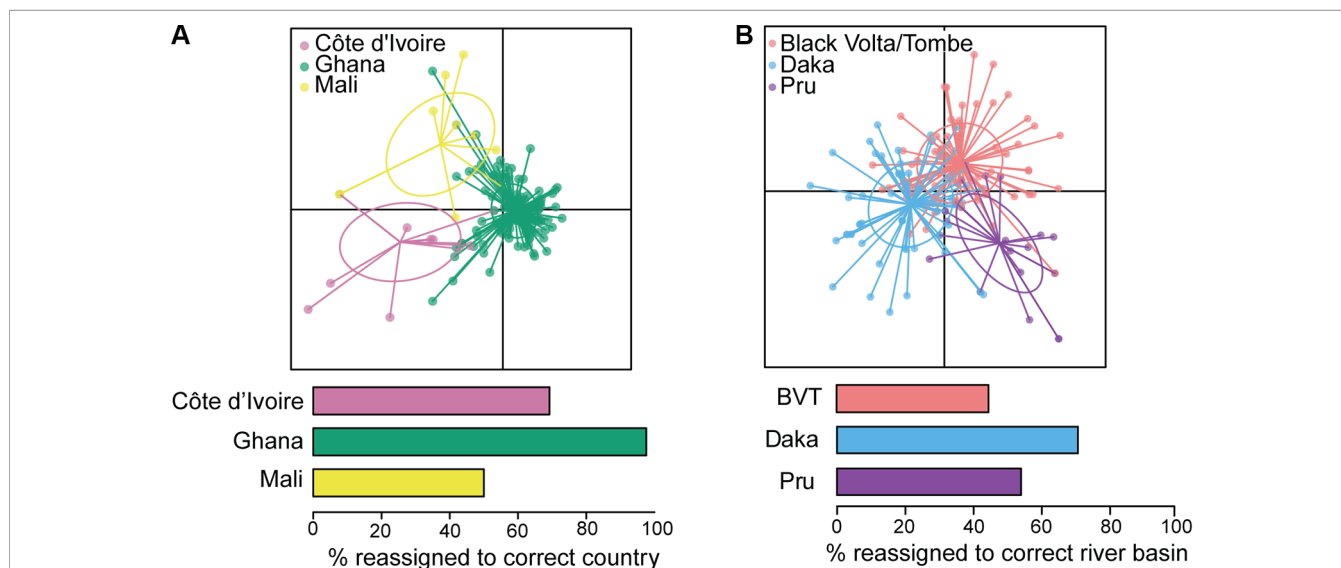
different species may influence transmission dynamics differently. Specific vector–parasite associations have been postulated (Duke, 1981; Basáñez et al., 2009). Blackflies from forest areas transmit more infective larvae of *O. volvulus* than flies prevalent in savannah areas (Cheke and Garms, 2013), and onchocerciasis-associated ocular damage is more common in savannah than in forest areas (Dadzie et al., 1990).

Because of this higher prevalence of onchocercal blindness in savannah compared to forest communities, early population genetics research focused on methods for differentiating "blinding savannah" and "non-blinding forest" parasite "strains." Nuclear genome variation in the O-150 repeat family differentiated *O. volvulus* collected from forest and savannah communities in West Africa and Nigeria (Erttmann et al., 1987; Erttmann et al., 1990; Zimmerman et al., 1992; Zimmerman et al., 1994a; Ogunrinade et al., 1999), but not Uganda or Sudan (Fischer et al., 1996; Higazi et al., 2001). DNA sequences derived from a ribosomal internal transcribed spacer (ITS-2) could not distinguish forest and savannah *O. volvulus* from West Africa (Morales-Hojas et al., 2007). Analyses of nuclear genome sequences quantified admixture between parasites collected from different forest and savannah sites: parasites can have a mix of ecotype-associated markers (Choi et al., 2016), indicating that parasites from forest and savannah can and do interbreed and their respective transmission zones overlap.

Reconstruction of mitochondrial genomes of a small number of geographically diverse adult *O. volvulus* (Choi et al., 2016) has pointed to greater genetic diversity than previously reported (Zimmerman et al., 1994b; Keddie et al., 1999). Crawford et al. (2019) analyzed whole mitochondrial genomes of ~150 parasites from West Africa and demonstrated that *O. volvulus* populations are extremely genetically diverse and that the historical population size was likely very large. More importantly in the context of this review, a subset of mitochondrial markers could distinguish *O. volvulus* sampled in Ghana from those sampled in Mali or Côte d'Ivoire (via discriminant analysis of principal components; **Figure 5A**). Nuclear sequence data differentiated worms collected in Ghana from those collected in Cameroon, indicating that current, ongoing interbreeding between parasites from these two countries is unlikely. Therefore, they are from different transmission zones (Doyle et al., 2017). At an even larger geographic scale, admixture analysis of genetic sequence data demonstrated that American parasite populations are derived from African populations: competent vectors existed in areas where infected people from Africa were forcibly brought to work as slaves (Zimmerman et al., 1994b; Unnasch and Williams, 2000; Choi et al., 2016).

Analysis of population structure over shorter distances within a single ecotype has been conducted for parasites from different river basins. Nuclear genetic differentiation was clear between individual parasites from the Mbam and Nkam river basins in Cameroon, suggesting no (or limited) parasite transmission between them, and that, consequently, the Mbam and Nkam river basins belong to two different transmission zones (Doyle et al., 2017). In practical terms, this means that decisions to stop treatment in one basin may not need to take into consideration ongoing transmission in the other. In contrast, both nuclear (Doyle et al., 2017) and mitochondrial genomes showed little genetic differentiation among parasites from three river basins across an east-west transect of ~250 km in Ghana: the Daka river basin on the eastern side of Lake Volta, the Pru river basin on the western side, and the Black Volta/Tombe river basin to the north (**Figure 5B**; Crawford et al., 2019). These results are consistent with weak population structure between the three basins, and significant past or current gene flow between them. The epidemiological interpretation of this population structure is

FIGURE 5 | Analyses of mitochondrial sequence data from West Africa. **(A)** Discriminant analysis of principal components (DAPC; Jombart et al., 2010) indicates that Ghana is distinct from Côte d'Ivoire and Mali. **(B)** DAPC indicates that three river basins within Ghana should be considered a single transmission zone. BVT, Black Volta/Tombe. From Crawford et al. (2019).

that transmission has occurred and likely continues to occur amongst these three river basins (Crawford et al., 2019). The Black Volta/Tombe, Pru, and Daka river basins should thus be considered a single transmission zone: cessation of treatment at any one location would require interruption of transmission throughout the zone to reduce risk of recrudescence.

The analyses of mitochondrial and nuclear variation presented here suggest that *O. volvulus* populations—and thus transmission zones—can cover large regions and different river basins. It remains to be seen to what extent these genetic estimates of transmission zones can be correlated with wind-facilitated vector movement, features of the local climate or landscape that impact fly prevalence because of breeding site availability, differences in transmission capacity of the local fly vector species, or factors impacting movement among regions by people such as low habitability, large distances, or seasonal migration of workers.

W. bancrofti Population Genetic Structure

Within populations of *W. bancrofti*, genetic diversity (i.e., the number of polymorphic sites, regardless of genetic background) and haplotype diversity (i.e., the number of unique, contiguous sequences) appear to be comparable to *O. volvulus* (de Souza et al., 2014; Choi et al., 2016; Small et al., 2016; Doyle et al., 2017; Crawford et al., 2019; Small et al., 2019). Diversity may be higher where transmission rates are higher (Blouin et al., 1992; Blouin et al., 1995), because population sizes are expected to be larger where there is more active transmission, and larger population sizes result in greater retention of genetic diversity over time. Studies in Papua New Guinea (PNG; Small et al., 2013), Ghana (de Souza et al., 2014), and India (Thangadurai et al., 2006; Hoti et al., 2008b; Mahakalkar et al., 2017) have found high levels of genetic variation among and within *W. bancrofti* populations, and suggested higher similarities between geographically proximate villages. Consistent with expectations that higher transmission rates (and consequently, larger parasite population size) may increase genetic diversity, Hoti et al. (2008b) detected the highest genetic variability in densely populated, urban areas.

Genetic differentiation has been associated with geographic distance and other barriers to movement of people and vectors in Ghana (de Souza et al., 2014), India (Kumar et al., 2002; Thangadurai et al., 2006; Hoti et al., 2007; Patra et al., 2007), and Nepal (Adhikari et al., 2015). For example, genetic fingerprinting detected divergence among strains on either side of the Western Ghats mountain range in India, likely driven by the geographic barrier (Thangadurai et al., 2006). Genetic differentiation between parasites from different vectors was identified in the Andaman and Nicobar Islands (Dhamodharan et al., 2008), and a nocturnally sub-periodic strain from Thailand was genetically distinguishable from a nocturnally periodic Myanmar strain (Nuchprayoon et al., 2007). These genetic differences could be driven by geographic distance reducing gene flow or a consequence of divergent phenotypic adaptation to enhance transmission by different vectors. A significant challenge with these studies is that many of them used RAPD assays, yielding data which are difficult to compare between studies and have uncertain reproducibility (Penner et al., 1993; Davin-Regli et al., 1995; Pérez et al., 1998).

Genomic sequencing of *W. bancrofti* within populations has been limited to 20 individual worms from PNG (Small et al., 2016; Small et al., 2019) and 27 from three geographically widespread communities (Haiti, Mali, and Kenya; Small et al., 2019). Small et al. (2019) analyzed all 47 genomes and found genetic differentiation among all countries sampled, suggesting migration is low or non-existent at this spatial scale. Further analyses supported ancestral movement of parasites between Africa and Haiti, likely driven by the forced imprisonment and transport of people from Western and Central Africa to the Americas (Small et al., 2019). These efforts indicate sufficient genetic variation in *W. bancrofti* to detect gene flow. Advances in whole-genome sequencing of *W. bancrofti*, such as those made by Small et al. (2016; 2019), using either infective larvae dissected from laboratory-reared mosquitoes or using selective whole-genome amplification of single microfilaria isolated from blood samples, may increase the feasibility and accuracy of using molecular sequence data to identify and track transmission of LF. In particular, advances in laboratory methods that could enable sequencing from materials commonly collected during TASs (e.g., from rapid diagnostic tests: Rao et al., 2006; microfilariae slides: Bisht et al., 2006; and dried blood spots: Kumar et al., 2019), would in turn increase availability of sequence data in terms of the number of individual parasites per community and the number of communities, allowing accurate delineation of transmission zones and identification of immigrant parasites.

PARASITE POPULATION GENOMICS FOR UNDERSTANDING GEOGRAPHIC AND TEMPORAL VARIABILITY IN DRUG RESPONSE

Conceptual Approach

One aim of population genomics is to identify genome variation associated with phenotypic variation in traits of interest. Because neither *O. volvulus* nor *W. bancrofti* can be cultured in the laboratory, this work requires correlations to be determined from worms collected from infected people and phenotyped for a trait (e.g., disease manifestation, variation in drug response) and then subsequently genotyped. For example, onchocerciasis disease severity and manifestations vary widely across Africa, with onchocerciasis-associated blindness more common in savannah regions (Remme et al., 1989; Dadzie et al., 1990), and severe onchodermatitis (Sowda, hyperreactive onchocerciasis) prevalent in Yemen and in one focus in Sudan (Mukhtar et al., 1998), but rare in other endemic areas (Al-Kubati et al., 2018). Similarly, the extent to which the level of skin microfilariae decreases after ivermectin treatment and the timing and extent of subsequent increases in skin microfilariae levels varies between individuals even in ivermectin-naive areas (Awadzi et al., 2014; Opoku et al., 2018).
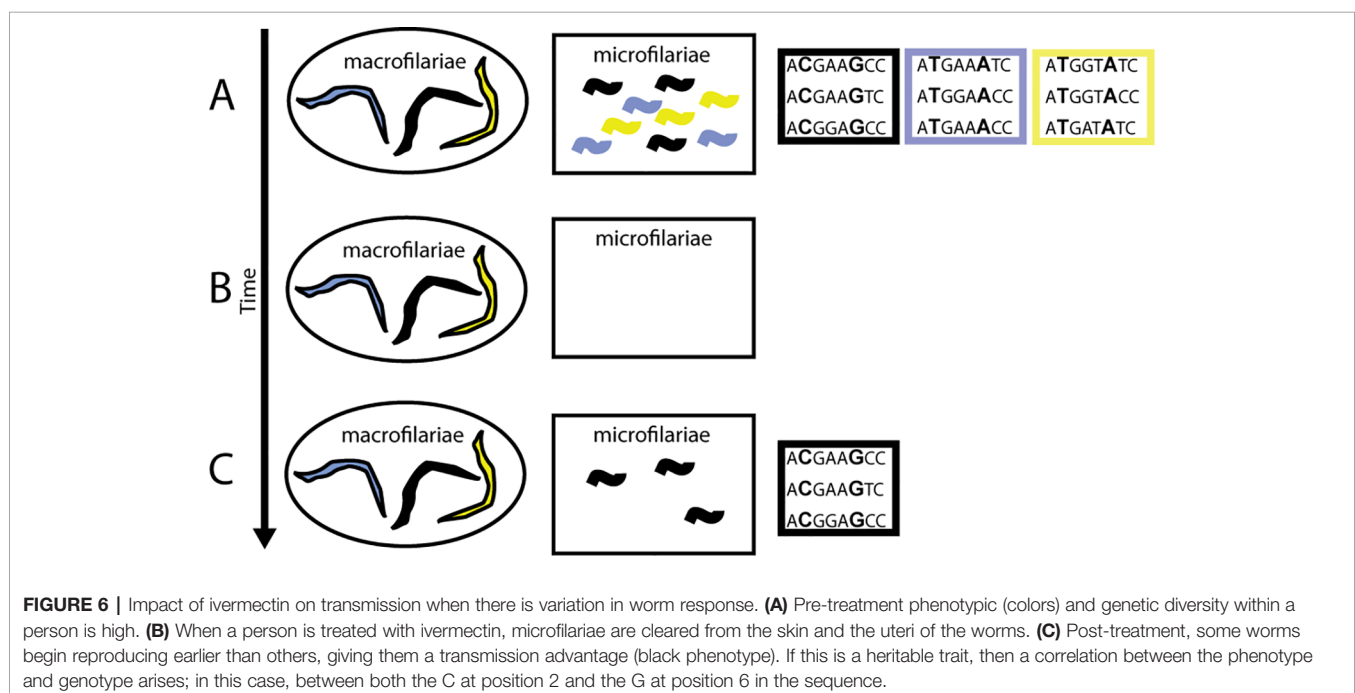
The challenge is determining whether there is a correlation between a well-identified phenotype (e.g., early resumption of

female worms' fertility after drug treatment) and a particular genotype (**Figure 6**). Identifying parasite population structure is a critical first step: it allows identification of genetic similarities and differences that arise because of degrees of relatedness between worms (interbreeding) and not because of genetic association with the parasites' phenotype. Such genetic variants should be excluded from analyses of correlations between phenotypic and genotypic variation. Equally important, and perhaps more challenging, is that the phenotype needs to be well defined. For example, when comparing differences in pathology, manifestations might depend on whether the person was infected as a child or as an adult, on variation in the longevity of microfilariae or adult worms in different geographic regions (e.g., forest and savannah worms; see Duke, 1993), or on the genetic background of the human host driving specific immunologic reactions (e.g., Sowda; see Hoerauf et al., 2002).

Defining Phenotypic Variation in Drug Response

The potential for the evolution of resistance to either macrocyclic lactones (such as ivermectin) or benzimidazoles (such as albendazole) in filarial nematodes infecting humans has been a concern from the start of MDA programs (e.g., WHO Expert Committee on Onchocerciasis Control 1993 and WHO, 1995), because of demonstrated evolution of resistance in parasites of domestic animals (Kaplan, 2004; Wolstenholme et al., 2004; Wolstenholme et al., 2015). However, identifying drug resistance in human-infecting filarial parasites that cannot be cultured in the laboratory is challenging, and quantifying genetic association with a phenotype is completely dependent on identifying the correct phenotype.

In *O. volvulus*, the phenomenon of "suboptimal response" to ivermectin (SOR) was identified during research into reasons for higher-than-expected prevalence of infection and skin microfilariae levels in areas under long-term MDAi in Ghana. SOR is defined as the resumption of microfilariae production by adult female worms following ivermectin treatment earlier than considered typical, resulting in detectable levels of skin microfilariae 80–90 days post-treatment and the presence of viable embryos and microfilariae in the uteri of female worms obtained through nodulectomy (Awadzi et al., 2004a; Awadzi et al., 2004b). Since reasons such as low ivermectin blood concentrations were excluded, Awadzi et al. attributed SOR to low susceptibility to the so-called "embryostatic" effect of ivermectin, resulting in developmental stages of microfilariae in the uteri 90 days after treatment (see also Grant, 1994). The progeny of these worms would be available for transmission earlier and for longer between rounds of MDAi than progeny of susceptible worms releasing microfilariae later (**Figure 6**). If low susceptibility to the "embryostatic" effect is heritable, the prevalence of SOR parasites may increase over time. While increasing prevalence of SOR worms is not expected to impact effectiveness of MDAi for control of onchocerciasis as a public health problem (because SOR does not change post-treatment disappearance of microfilariae from the skin—the "microfilaricidal effect"), it would negatively impact elimination, which depends on progressively lower skin microfilariae available for transmission. Since the initial description of SOR, further investigations of SOR in Ghanaian and Cameroonian areas with different ivermectin treatment histories (e.g., Osei-Atweneboana et al., 2007; Osei-Atweneboana et al., 2011; Pion et al., 2013; Nana-Djeunga et al., 2014) have shown that SOR parasites are detectable even

FIGURE 6 | Impact of ivermectin on transmission when there is variation in worm response. **(A)** Pre-treatment phenotypic (colors) and genetic diversity within a person is high. **(B)** When a person is treated with ivermectin, microfilariae are cleared from the skin and the uteri of the worms. **(C)** Post-treatment, some worms begin reproducing earlier than others, giving them a transmission advantage (black phenotype). If this is a heritable trait, then a correlation between the phenotype and genotype arises; in this case, between both the C at position 2 and the G at position 6 in the sequence.

where MDAi occurs biannually (Frempong et al., 2016). Modeling showed large inter-individual variability in skin microfilariae repopulation rates among individuals from Ghana, Liberia, Sierra Leone, and Guatemala treated one to five times with ivermectin (Churcher et al., 2009). None of the individuals from Ghanaian communities who received 10–19 rounds of ivermectin treatment (Osei-Atweneboana et al., 2007) had a higher repopulation rate higher than the highest observed in the control data set (skin microfilariae repopulation rate after the first ivermectin treatment among individuals from Ghana, Liberia, Sierra Leone, and Guatemala), but analysis by village showed significantly higher repopulation rate for some villages and for one previously untreated village in the same area (Churcher et al., 2009) This supports the conclusion that SOR parasites are present in ivermectin-naive communities and may increase *via* selection by ivermectin treatment. Furthermore, variability of response to ivermectin measured by skin microfilariae levels over time after a single dose of ivermectin was found among individuals from ivermectin-naive areas in Ghana, Liberia, and the Democratic Republic of Congo: the skin microfilariae kinetics in some of these individuals met the SOR criteria used in other studies (Awadzi et al., 2004a; Awadzi et al., 2004b; Opoku et al., 2018). It is unknown whether these ivermectin-naive areas belong to the same transmission zones as other nearby areas with ongoing MDAi. Together, these data support the conclusion that *O. volvulus* populations consist of worms with variable susceptibility to the embryostatic effect of ivermectin. Based on these data, the phenotype for genomic analysis is the reproductive status of individual female worms collected from people 80–90 days post-ivermectin treatment: female worms that respond well to ivermectin should not have post-fertilization embryonic stages, while those considered SOR have early resumption of fecundity, and thus already have microfilariae in uteri (Osei-Atweneboana et al., 2011; Nana-Djeunga et al., 2014; Doyle et al., 2017).

In *W. bancrofti,* reasons underlying continued persistence of transmission are challenging to determine. So far persistence has not yet been associated with heritable variability in parasite drug response. Density-dependent effects on vector response to infection and vector competence play a role in continuous, low-level transmission, depending on the vector species: iL3 development is proportionally higher at lower microfilariae densities in some vector species, while in other vectors, fewer iL3 develop when many microfilariae are ingested (Pichon, 2002; reviewed in Graves et al., 2016). These differences change transmission dynamics during MDA in ways highly dependent on specific vector species that may favor greater vector efficiency when microfilaremia is low.

Defining Correlations Between Phenotypic and Genotypic Variation for Drug Response

Several point mutations in the target of albendazole, β-tubulin, lead to drug resistance in nematodes (Kwa et al., 1994; Grant and Mascord, 1996; Silvestre and Cabaret, 2002; Silvestre and Humbert, 2002), and at least one of these has been identified

in *W. bancrofti* (Schwab et al., 2005). Increases in alleles associated with benzimidazole resistance have been detected after MDA (Schwab et al., 2005), and modeling suggests that further MDA treatments would lead to selection *via* an increase in the frequency of these alleles (Schwab et al., 2006; Schwab et al., 2007). However, analysis of natural variation in response to benzimidazole anthelmintics in the free-living nematode *Caenorhabditis elegans* has shown that many other loci contribute to and modify variation at the β-tubulin locus in this species (Hahnel et al., 2018), and it is clear from analysis on benzimidazole resistance in veterinary helminths that β-tubulin–mediated resistance is also subject to modification by as yet unidentified loci (Doyle and Cotton, 2019).

With respect to DEC, this drug is not 100% effective against LF (Eberhard et al., 1988; Eberhard et al., 1991), suggesting that variation in DEC susceptibility exists and could lead to the development and spread of drug resistance if poor susceptibility is both heritable and mating and recombination allow for assortment among alleles (Grant, 1994; McCarthy, 2005; Cobo, 2016).

The view that has emerged from single-candidate gene association studies in helminths is that response to macrocyclic lactones such as ivermectin (a glutamate-gated chloride channel agonist) is a complex and highly variable phenotype, affected by multiple genes across the genome (e.g., Dent et al., 2000; Ardelli et al., 2009; Yan et al., 2012; Ardelli and Prichard, 2013; Menez et al., 2016), with poor repeatability between studies, strains, and species. A single-candidate gene approach has been applied to ivermectin SOR in *O. volvulus* by sequencing genes that might contribute to ivermectin resistance in other nematodes, and comparing differences between communities that have never received ivermectin and those that have, or between microfilarial pools collected before and after treatment (Ardelli and Prichard, 2004; Ardelli et al., 2005; Eng and Prichard, 2005; Eng et al., 2006; Ardelli and Prichard, 2007; Osei-Atweneboana et al., 2012). If a gene contributes to ivermectin response, the expectation is that alleles that decrease sensitivity to ivermectin would be more common in worms from areas that have had many rounds of MDAi than in worms from ivermectin-naive areas, or that they would be more common in microfilariae collected shortly after ivermectin treatment than before treatment, or would be more common in adult females that have resumed reproduction within 80–90 days of treatment (the SOR phenotype). Genes identified using this approach include β-tubulin (Eng and Prichard, 2005; Bourguinat et al., 2006; Eng et al., 2006; Bourguinat et al., 2007; Osei-Atweneboana et al., 2012) and members of the ATP-binding cassette transporter family (Ardelli and Prichard, 2004; Ardelli et al., 2005; Eng and Prichard, 2005; Ardelli et al., 2006a; Ardelli et al., 2006b; Ardelli and Prichard, 2007).

However, this approach faces one of the biggest challenges in detecting selection: the frequency of an allele can increase or decrease due to random chance rather than selection. In particular, if a population decreases in size, as might be expected as transmission decreases during successive rounds of MDA, the chance that changes in allele frequencies are stochastic

rather than deterministic increases: the genetic diversity overall decreases (e.g., *Schistosoma mansoni*: Coeli et al., 2013), but the fate of individual alleles is unpredictable, and some may increase due to genetic drift caused by population decline. Another challenge is that, prior to treatment, parasite populations can be quite large (i.e., they have a high effective population size; Blouin et al., 1995; Blouin et al., 1999; Zhou et al., 2013; Rodelsperger et al., 2014; Jan et al., 2016), which means they have ample genetic variation, so genetic variation contributing to a poor response phenotype may be present at different frequencies and at different loci in different parasite populations (e.g., *Haemonchus contortus*, Gilleard and Redman, 2016). Furthermore, once MDA begins, mating and recombination could create new combinations of alleles from different genes, which may also decrease response to ivermectin. In other words, good responders may be carriers of alleles that, in the right genomic background, could contribute to poor response in drug naive populations, and vice versa. Selection then acts on "standing" genetic variation that was present in the population before the selection pressure began. This mechanism for selection has been called a "soft sweep" (e.g., Hermisson and Pennings, 2005; Przeworski et al., 2005; Pritchard and Di Rienzo, 2010; Ferrer-Admetlla et al., 2014; see also Doyle et al., 2017) and is notoriously difficult to detect because (1) each variant alone could have a weak effect on phenotype, and thus difficult to distinguish statistically from background noise (random variation), and (2) many techniques for detecting selection rely on correlated, linked variants being physically near to the causative mutation, but in a soft sweep, this linkage is often weak (Pennings and Hermisson, 2006). The technical and conceptual weaknesses of candidate gene approaches specifically in the context of anthelmintic response was reviewed recently (Doyle and Cotton, 2019).

Genome-wide approaches are, therefore, required. Doyle et al. (2017) examined female *O. volvulus* macrofilariae from Ghana and Cameroon that were "good responders" and "suboptimal responders" to the embryostatic effect of ivermectin, determined by the absence or presence of stretched microfilariae in the uteri. Females were pooled by response phenotype, sequenced, and genetic variants correlated with phenotype identified. Genetic variants clustered in 31 quantitative trait loci (QTLs) with genes for molecular pathways involved in neurotransmission, development, and stress response. None of these QTLs contained genes previously proposed by single candidate association studies. In addition, some loci associated with SOR to ivermectin in worms from Ghana were different from loci associated with SOR in worms from Cameroon. They concluded that ivermectin response is a polygenic, quantitative trait, subject to soft sweeps of pre-existing QTLs, rather "hard selection" on rare, resistance-conferring mutations. Since populations possess different standing variation, selection may act on different loci in different populations (Doyle et al., 2017). Genome-wide studies of veterinary nematodes generally support the "soft sweep QTL" model for response to macrocyclic lactones (Bourguinat et al., 2015; Redman et al., 2015; Choi et al., 2017; Doyle et al., 2019), and likewise have not identified any of the candidate genes described above.

This "soft sweep QTL" model that has emerged from genome-wide association studies of ivermectin response leads to the following predictions for control and elimination of human filariases:

(1) Substantial natural variation in ivermectin (and perhaps benzimidazole or DEC) response may exist within and between populations at the beginning of MDA. Therefore, in some populations, MDA could reduce transmission more efficiently than in others, ultimately leading to elimination success in some populations but not others. Genetic markers that predict drug response can first be defined (as has been done for *D. immitis* in the USA; Bourguinat et al., 2015; Bourguinat et al., 2017; Ballesteros et al., 2018) and then used to identify "at-risk" populations in which the initial SOR allele frequency is high enough to prevent elimination by MDA alone.

(2) Selection for increased SOR allele frequency will likely occur *via* soft sweeps, making detection of underlying genetic changes more challenging. That said, a small panel of SNP alleles can predict treatment response in *D. immitis* (with the caveat that the population structure of *D. immitis* in these studies is not reported; Bourguinat et al., 2015; Bourguinat et al., 2017; Ballesteros et al., 2018).

(3) Ivermectin response appears to be a highly complex phenotype, and changes in the molecular targets of ivermectin may not be the cause of SOR. For example, selection could favor worms with a faster rate of embryonic development, or resumption of microfilariae production might be driven not by female recovery from an embryostatic effect, but by recovery of males and thus earlier insemination, or by recovery of one or both sexes from an effect that impairs mating. The possibility that the embryostatic effect of ivermectin is a male- rather than female-mediated phenomenon has not been tested in *O. volvulus* or LF, but suppression of embryogenesis in *D. immitis* appears to be due to effects of macrocyclic lactones on the reproductive competency of male and not female worms (Lok et al., 1995). This has implications for how the drug response phenotype should be defined.

(4) It is now clear that variation in drug response is heritable. If SOR alleles are initially present and selected for, even in only some communities undergoing MDAi, SOR could spread throughout the whole transmission zone (including to hypoendemic areas not initially included in MDAi). An important secondary advantage to identifying transmission zones, then, is that the potential risks and time frame of the spread of a SOR genotype to connected communities could be quantified.

DEVELOPMENT OF POPULATION GENOMICS–BASED TOOLS TO INFORM ELIMINATION PROGRAM DECISIONS

A wide range of DNA-based molecular tools have been developed for detecting the presence of parasites in vectors and

the human host using skin, blood, plasma, and even urine (e.g., Zimmerman et al., 1993; Rao et al., 2006; Rodriguez-Pérez et al., 2006; Hoti et al., 2008a; Gass et al., 2012; Alhassan et al., 2015; Owusu et al., 2015; Lau et al., 2016). The addition of information on transmission gained from genomic analyses of parasites/vectors to data on prevalence and intensity could significantly improve the basis for elimination program decisions on whether to stop or continue interventions, determine causes for recrudescence post-intervention and thus appropriate subsequent interventions, and determine optimal post-intervention surveillance frequencies commensurate with the risk of resurgence based on a cost–risk evaluation.

Methods to obtain and analyze genome sequences require training and equipment not available to many laboratories in endemic countries. However, once the required basic genome research has been done, fast and relatively inexpensive methods can be developed suitable for use in at least one laboratory in an endemic country, possibly supported by one or more "reference laboratories."

Tools for Defining Transmission Zones and Identification of Migrants

The "genetic signatures" of an interbreeding parasite population constitute markers that allow delineation of transmission zones as "the natural ecological and epidemiological unit for interventions" for onchocerciasis or the implementation and evaluation units for LF. In conjunction with transmission models that allow modeling the impact of immigrant parasites on long-term infection prevalence, the size of these zones can inform cost–risk analysis for stop-treatment decisions as well as when and where to conduct entomological evaluations or TAS, dependent on the risk of resurgence that programs are willing to accept.

The advantage of DNA-based transmission zone markers is that they can be designed to be highly specific. The first step is identifying variant sites informative for population structure; i.e., are polymorphic and more common or fixed (and thus have different frequency-based likelihoods) in genomes of parasites from one area compared to another. For example, Crawford et al. (2019) identified mitochondrial polymorphic sites that could differentiate *O. volvulus* collected in Ghana from countries farther west with high statistical likelihood. Once such variant sites have been identified for specific geographic areas, parasite genomes from other areas need to be analyzed to determine whether the same variants are informative, or whether additional sites are necessary to infer the extent of interbreeding. Once analysis is completed for parasites from across their distribution, a minimum set of polymorphic sites can be defined to predict the transmission zone of origin for any given worm.

This process may sound daunting, but methodology developed over the past years is efficient and amenable to further optimization. Skin snips or infected/infective *Simulium* or *O. volvulus* adults from many areas in Africa are available in the ESPEN laboratory in Ouagadougou, national laboratories, and research centers. LF parasites, primarily on slides of blood films and in dried blood spots, are available from endemic areas in national laboratories and research centers.

Developing inexpensive methodologies for routine use in each endemic country is the next challenge. High resolution melt (HRM) analysis has been proposed as a species identification tool for *Onchocerca* spp. collected from blackflies (Doyle et al., 2016) and could potentially be extended as a genotyping tool. The advantage to using vectors is that, in the case of *Simulium* spp., vector collection is already an integral part of the process for stop-treatment decisions, invasive sampling of people would not be required, and (if sampling strategies are well designed) vectors would include a representative sample of parasites from many different human hosts.

Tools for Monitoring Drug Susceptibility

Ideally, control/elimination programs would have reference data on pre-MDA/early-MDA variability of parasite drug response and an easy-to-use, cost-effective, and non-invasive tool for monitoring changes in response phenotype frequency. Detecting changes while the frequency of poor responder alleles is low would allow alternative control/elimination strategies to be put in place (Grant, 2000) and enable evidence-based decisions on monitoring strategies. Some moves toward these ends have been initiated (e.g., Ardelli and Prichard, 2004; Eng and Prichard, 2005; Bourguinat et al., 2006; Bourguinat et al., 2008; Hoti et al., 2009) but have yet to result in practical application. Research for methods development for monitoring potential emergence of resistance to ivermectin in *O. volvulus* were initiated by the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR) and OCP in the early 1990s (WHO Expert Committee of Onchocerciasis Control (1993) and WHO, 1995; UNDP/World Bank/WHO Special Programme for Research & Training in Tropical Diseases, 1997). TDR also initiated characterization of parasite drug response variability before or early during MDA (Walker et al., 2016; Halder et al., 2017), research which continues with Wellcome Trust funding leveraged with this initial TDR investment. Alternative strategies for onchocerciasis elimination include increased MDAi frequency, complementary vector control, test-and-treat strategies, and new drugs (African Programme for Onchocerciasis Control, 2015; Boussinesq et al., 2018; Verver et al., 2018; Horstick and Runge-Ranzinger, 2018).

Currently, screening communities for *O. volvulus* with low ivermectin susceptibility requires either collection of skin snips before and after ivermectin treatment to determine skin microfilariae levels, or post-treatment nodulectomies to determine the reproductive status of individual female worms. Genetic analysis of microfilariae collected *via* skin snips at a single post-treatment time point, or larvae in vectors, could significantly improve monitoring capacity. While the molecular mechanisms that contribute to variation in drug response are of significant interest, identifying anonymous genetic variants associated with drug response phenotype is sufficient for development of tools for monitoring, even while research into the mechanisms is ongoing.

Genetic variation associated with ivermectin response in *O. volvulus* has been identified in parasites from Ghana and Cameroon (Doyle et al., 2017). These results have not yet been validated with

single-worm genomic analysis, and whether the response-associated genetic variants identified are shared across populations from other areas in Africa is untested. Compared to development of transmission zone markers, development of ivermectin response markers is much more challenging, since large numbers of parasites with known response phenotype from different geographic areas are needed to identify, validate, and test the predictive value of genetic variants associated with response phenotype.

Identification of a panel of relatively few variants (e.g., < 50) would pave the way for developing a simple PCR-based or loop-mediated isothermal amplification (LAMP) surveillance tool for use with parasites in vectors or skin snips by endemic country laboratories. This core set of variants would define the QTLs that are both necessary and sufficient for SOR to develop in *O. volvulus* and may, therefore, shed light on the as yet unexplained mechanism by which ivermectin exerts its embryostatic effect.

MDA for *W. bancrofti* uses albendazole, alone or in combination with DEC and/or ivermectin. A PCR-based assay for one single nucleotide polymorphism that leads to resistance to benzimidazoles has been developed, so that monitoring for allele changes in treated communities can occur (Hoti et al., 2009). However, while variation in response to drug therapies for LF has been reported (e.g., Eberhard et al., 1988; Eberhard et al., 1991), to date a clear definition of an LF SOR has not been reported (Irvine et al., 2017). A clear, and preferably quantitative, definition of poor response is an essential prerequisite for development of response genetic markers in LF.

Models Incorporating Population Genetics to Estimate Risk of Recrudescence From Migration or Poor Drug Response

There is a long history of the use of onchocerciasis and LF transmission models such as EPIONCHO and ONCHOSIM (reviewed in Stolk et al., 2015; Basáñez et al., 2016) and EPIFIL, LYMFASIM, and TRANSFIL (reviewed in Smith et al., 2017) to inform control program decisions. These models differ in whether they are deterministic or stochastic and whether they are population- or individual-based, and can be parameterized to match the characteristics of a particular community, including endemicity, variation in biting rates, or treatment coverage, such that effects of different treatment strategies on transmission can be explored. However, these programs currently cannot model movement of parasites from one geographic area to another (via people or vectors), spatial heterogeneity of endemicity or elimination program implementation, or transmission of parasites with variable drug susceptibility. Thus, although these models have been useful in estimating declines of parasite infection prevalence and intensity over a control campaign, they do not realistically model variation in parasite drug susceptibility, human/vector movement among areas, or long-term population processes that could impede reaching elimination thresholds or contribute to medium- to long-term recrudescence after MDA discontinuation.

To address these limitations, TDR funded extension of transmission models for preventive chemotherapy-controlled diseases. Coffeng et al. (2017) incorporated evolution of changes

in drug susceptibility in *O. volvulus* (and other parasites) into a stochastic, individual-based model, and McCulloch et al. (2017) adapted an infection-intensity model, EPIONCHO, into a patch framework for onchocerciasis, incorporating differences in endemicity and treatment between two adjacent geographic areas, or "patches," and people and vector migration between patches. When simulating migration of seasonal workers from a patch where transmission is ongoing to a patch where MDAi had interrupted transmission, modest levels of migration (10% of the adult population move for between 1 and 12 months) were sufficient to re-establish transmission with a lag time of 5–10 years, and pre-MDA prevalence was reached within 30–50 years in the absence of renewed MDA.

Data from Burkina Faso showed a significant delay between cessation of interventions and recrudescence (Koala et al., 2017; Koala et al., 2019), suggesting that the predicted time frames for recrudescence modeled by McCulloch et al. (2017) are realistic. Such delays are not unexpected considering that parasite invasion may initiate transmission at a low level, and transmission takes time to become detectable. A model incorporating population genetic structure data, allowing simulation of vector and human movement–mediated parasite transmission between different geographic areas, could provide estimates of the risk and time frame of resurgence in one area where treatment cessation is planned while transmission is above stop-treatment thresholds elsewhere. National programs could obtain a measure of risk for stopping interventions given migration and an estimated time frame for resurgence, decide whether to intensify interventions in areas that are potential source areas of invasion, and plan cost-effective, post-MDA surveillance strategies.

A model parameterized for a specific region can be informative for elimination programs by exploring how variation in treatment coverage, changes in human demographic and travel patterns, vector control and migration patterns, or other factors will affect progress toward and post-intervention sustainability of elimination within different areas in a region. Such a spatially explicit, agent-based modeling framework was developed to explore LF transmission in the Samoan Islands (GEOFIL; Xu et al., 2018a). This type of model incorporates detailed information about individuals within the community and their connectivity, and simulates changes in characteristics such as age, physical location, and infection status. GEOFIL predicts spatial variation in LF infection prevalence across American Samoa over space and time. The model can contribute toward understanding transmission dynamics in areas of high infection prevalence (hot spots) which can help determine effective surveillance strategies for identifying hot spots, and how best to manage them. Incorporation of information from genomic analyses could enhance the utility of such models for elimination decisions. An agent-based framework does require significant data on individual people, information not available for many (or most) communities, and it will be valuable to see how lessons from a well-characterized system can be applied more broadly to areas where LF has re-emerged post-MDA.

CONCLUSIONS AND RECOMMENDATIONS

Population genomic analyses of *O. volvulus* and *W. bancrofti* show high levels of genetic variation consistent with historically large population sizes (> 10^5). Genetic variation is spatially structured; parasite populations are genetically distinct, indicating that they are not interbreeding. For *O. volvulus*, genomic data from West Africa indicate that the geographic areas across which parasites interbreed—transmission zones—can cover hundreds of kilometers and extend beyond the boundaries of a "river basin, or a major section of a river basin" (African Programme for Onchocerciasis Control/World Health Organization, 2010). This large-scale structure is reflected in the pattern of onchocerciasis endemicity in Africa, with a mosaic of hypo-, meso-, and hyperendemic areas (**Figure 3**). For *W. bancrofti*, transmission zones are larger than expected based on the focal nature of many mosquito vector populations, consistent with the increasingly recognized role that movement of infected people plays. In epidemiological terms, genetic data indicate that significant transmission must have occurred (and likely to still be occurring) over tens to hundreds of kilometers.

We propose that long-range transmission *via* infected people and/or infective vectors has epidemiological significance for elimination programs that it does not have for control programs. The latter plan to continue treatment indefinitely, and consequently, the effect of occasional or continual long-range transmission, is suppressed by continuing MDA. In contrast, elimination programs stop treatment once relevant criteria have been met within a defined area, assuming transmission has been interrupted permanently. Without MDA, occasional or low but continual parasite transmission from outside an intervention area may result in recrudescence. Population genetic analyses can predict whether the potential for gene flow is epidemiologically significant. Consequently, elimination programs need to define the boundaries of the entire extended geographic area over which parasite transmission occurs and continue treatment until transmission is interrupted over the entire area. From an operational perspective, population genomic statistics do not differentiate between long-range transmission due to movement of vectors and that due to movement of people, and are thus informative independent of the cause of parasite migration.

What research and development are required to allow elimination programs to take advantage of population genomics? One clear need is gathering more (and better quality) sequence data across a greater geographical range, based on improved methods which allow sequencing from parasite material already available or routinely collected (blackflies or skin snips for onchocerciasis, and mosquitoes or blood samples for LF). Testing and development of analytical methods for quantifying parasite migration based on genetic, entomological, and epidemiological data, and their optimal application to transmission zone delineation, is critical. Tools not requiring specialized laboratory equipment must be developed for use by national programs to include in MDA and stop-MDA evaluations. These tools would allow investigation of the source of infections/transmission identified during post-treatment surveillance, facilitate monitoring for changes in the frequency of parasites with SOR to MDA, and promote planning cost-effective solutions. Comparable population genomic data for vectors would be a useful adjunct to differentiate between vector and human movement as sources of long-range transmission to inform appropriate interventions. Current transmission models need to be enhanced to allow modeling of areas with different endemicity, treatment history, progress toward elimination, heritable drug resistance, and parasite gene flow, to better estimate the risk and time frame of recrudescence when treatment is to be stopped in one area while transmission is continuing in another. Finally, the research and development outputs can improve procedures and criteria for cessation of treatment, including estimating risks of recrudescence to inform cost–risk assessment-based stop-treatment decisions and appropriate post-treatment monitoring strategies.

AUTHOR CONTRIBUTIONS

SH, KC, AK, PG, and WG wrote the first draft. SH, AK, KC, PG, MB, CL, DB, and WG critically reviewed, edited, and approved the final version of the manuscript. The authors alone are responsible for the views expressed which do not necessarily represent the views, decisions, or policies of the institutions with which the authors are affiliated.

REFERENCES

Adhikari, R. K., Sherchand, J. B., Mishra, S. R., Ranabhat, K., and Wagle, R. R. (2015). Awareness and coverage of mass drug administration for elimination of lymphatic filariasis: a community based cross sectional study in Nepal. *J. Community Health* 40 (1), 34–40. doi: 10.1007/s10900-014-9891-1

African Programme for Onchocerciasis Control and World Health Organization (2010). *Conceptual and operational framework of onchocerciasis elimination with ivermectin treatment* (Geneva: World Health Organization).

African Programme for Onchocerciasis Control and World Health Organization (2013). *The World Health Organization year 2013 progress report* (Geneva: World Health Organization).

African Programme for Onchocerciasis Control (2012). *Final communiqué of the 18th session of the Joint Action Forum (JAF18) of APOC* (Geneva: World Health Organization).

African Programme for Onchocerciasis Control (2015). *The WHO African Programme for Onchocerciasis Control final evaluation report, October 2015* (Geneva: World Health Organization).

Alhassan, A., Li, Z., Poole, C. B., and Carlow, C. K. (2015). Expanding the MDx toolbox for filarial diagnosis and surveillance. *Trends Parasit.* 31 (8), 391–400. doi: 10.1016/j.pt.2015.04.006

Al-Kubati, A. S., Mackenzie, C. D., Boakye, D., Al-Qubati, Y., Al-Samie, A. R., Awad, I. E., et al. (2018). Onchocerciasis in Yemen: moving forward towards an elimination program. *Int. Health* 10(suppl_1), i89–i96. doi: 10.1093/inthealth/ihx055

Ardelli, B. F., and Prichard, R. K. (2004). Identification of variant ABC-transporter genes among *Onchocerca volvulus* collected from ivermectin-treated and untreated patients in Ghana, West Africa. *Ann. Trop. Med. Parasitol* 98 (4), 371–384. doi: 10.1179/000349804225003415

Ardelli, B. F., and Prichard, R. K. (2007). Reduced genetic variation of an *Onchocerca volvulus* ABC transporter gene following treatment with ivermectin. *Trans. R Soc. Trop. Med. Hyg.* 101 (12), 1223–1232. doi: 10.1016/j.trstmh.2005.03.019

Ardelli, B. F., and Prichard, R. K. (2013). Inhibition of P-glycoprotein enhances sensitivity of *Caenorhabditis elegans* to ivermectin. *Vet. Parasitol* 191 (3-4), 264–275. doi: 10.1016/j.vetpar.2012.09.021

Ardelli, B. F., Guerriero, S. B., and Prichard, R. K. (2005). Genomic organization and effects of ivermectin selection on *Onchocerca volvulus* P-glycoprotein. *Mol. Biochem. Parasitol.* 143 (1), 58–66. doi: 10.1016/j.molbiopara.2005.05.006

Ardelli, B. F., Guerriero, S. B., and Prichard, R. K. (2006a). Characterization of a half-size ATP-binding cassette transporter gene which may be a useful marker for ivermectin selection in *Onchocerca volvulus.* *Mol. Biochem. Parasitol.* 145 (1), 94–100. doi: 10.1016/j.molbiopara.2005.09.011

Ardelli, B. F., Guerriero, S. B., and Prichard, R. K. (2006b). Ivermectin imposes selection pressure on P-glycoprotein from *Onchocerca volvulus*: linkage disequilibrium and genotype diversity. *Parasitology* 132 (Pt 3), 375–386. doi: 10.1017/S0031182005008991

Ardelli, B. F., Stitt, L. E., Tompkins, J. B., and Prichard, R. K. (2009). A comparison of the effects of ivermectin and moxidectin on the nematode *Caenorhabditis elegans.* *Vet. Parasitol.* 165 (1-2), 96–108. doi: 10.1016/j.vetpar.2009.06.043

Ariey, F., Duchemin, J. B., and Robert, V. (2003). Metapopulation concepts applied to falciparum malaria and their impacts on the emergence and spread of chloroquine resistance. *Infect. Genet. Evol.* 2, 185–192. doi: 10.1016/S1567-1348(02)00099-0

Awadzi, K., Attah, S. K., Addy, E. T., Opoku, N. O., Quartey, B. T., Lazdins-Helds, J. K., et al. (2004a). Thirty-month follow-up of sub-optimal responders to multiple treatments with ivermectin, in two onchocerciasis-endemic foci in Ghana. *Ann. Trop. Med. Parasitol.* 98 (4), 359–370. doi: 10.1179/000349804225003442

Awadzi, K., Boakye, D. A., Edwards, G., Opoku, N. O., Attah, S. K., Osei-Atweneboana, M. Y., et al. (2004b). An investigation of persistent microfilaridermias despite multiple treatments with ivermectin, in two onchocerciasis-endemic foci in Ghana. *Ann. Trop. Med. Parasitol.* 98 (3), 231–249. doi: 10.1179/000349804225003253

Awadzi, K., Opoku, N. O., Attah, S. K., Lazdins-Helds, J., and Kuesel, A. C. (2014). A randomized, single-ascending-dose, ivermectin-controlled, double-blind study of moxidectin in *Onchocerca volvulus* infection. *PloS Negl. Trop. Dis.* 8 (6), e2953. doi: 10.1371/journal.pntd.0002953

Aydemir, O., Janko, M., Hathaway, N. J., Verity, R., Mwandagalirwa, M. K., Tshefu, A. K., et al. (2018). Drug-resistance and population structure of *Plasmodium falciparum* across the Democratic Republic of Congo using high-throughput molecular inversion probes. *J. Infect. Dis.* 218 (6), 946–955. doi: 10.1093/infdis/jiy223

Bakajika, D., Senyonjo, L., Enyong, P., Oye, J., Biholong, B., Elhassan, E., et al. (2018). On-going transmission of human onchocerciasis in the Massangam health district in the West Region of Cameroon: better understanding transmission dynamics to inform changes in programmatic interventions. *PloS Negl. Trop. Dis.* 12 (11), e0006904. doi: 10.1371/journal.pntd.0006904

Baker, R. H., Guillet, P., Seketeli, A., Poudiougo, P., Boakye, D., Wilson, M. D., et al. (1990). Progress in controlling the reinvasion of windborne vectors into the western area of the Onchocerciasis Control Programme in West Africa. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 328, 731–747. doi: 10.1098/rstb.19900141 discussion 747-750.

Ballesteros, C., Pulaski, C. N., Bourguinat, C., Keller, K., Prichard, R. K., and Geary, T. G. (2018). Clinical validation of molecular markers of macrocyclic lactone resistance in *Dirofilaria immitis.* *Int. J. Parasitol. Drugs Drug Resist.* 8 (3), 596–606. doi: 10.1016/j.ijpddr.2018.06.006

Barrett, L. G., Thrall, P. H., Burdon, J. J., and Linde, C. C. (2008). Life history determines genetic structure and evolutionary potential of host-parasite interactions. *Trends Ecol. Evol.* 23 (12), 678–685. doi: 10.1016/j.tree.2008.06.017

Basáñez, M. G., Churcher, T. S., and Grillet, M. E. (2009). *Onchocerca-Simulium* interactions and the population and evolutionary biology of *Onchocerca volvulus.* *Adv. Parasitol.* 68, 263–313. doi: 10.1016/S0065-308X(08)00611-8

Basáñez, M. G., Walker, M., Turner, H. C., Coffeng, L. E., de Vlas, S. J., and Stolk, W. A. (2016). River blindness: mathematical models for control and elimination. *Adv. Parasitol.* 94, 247–341. doi: 10.1016/bs.apar.2016.08.003

Bhumiratana, A., Pechgit, P., Koyadun, S., Siriaut, C., and Yongyuth, P. (2010). Imported bancroftian filariasis: diethylcarbamazine response and benzimidazole susceptibility of *Wuchereria bancrofti* in dynamic cross-border migrant population targeted by the National Program to Eliminate Lymphatic Filariasis in South Thailand. *Acta Trop.* 113 (2), 121–128. doi: 10.1016/j.actatropica.2009.10.004

Bhumiratana, A., Intarapuk, A., Koyadun, S., Maneekan, P., and Sorosjinda-Nunthawarasilp, P. (2013). Current bancroftian filariasis elimination on Thailand-Myanmar border: public health challenges toward postgenomic MDA evaluation. *ISRN Trop. Med.* 2013, 857935. doi: 10.1155/2013/857935

Bisht, R., Hoti, S. L., Thangadurai, R., and Das, P. K. (2006). Isolation of *Wuchereria bancrofti* microfilariae from archived stained blood slides for use in genetic studies and amplification of parasite and endosymbiont genes. *Acta Trop.* 99 (1), 1–5. doi: 10.1016/j.actatropica.2005.12.009

Blouin, M. S., Dame, J. B., Tarrant, C. A., and Courtney, C. H. (1992). Unusual population genetics of a parasitic nematode: mtDNA variation within and among populations. *Evolution* 46 (2), 470–476. doi: 10.1111/j.1558-5646.1992.tb02052.x

Blouin, M. S., Yowell, C. A., Courtney, C. H., and Dame, J. B. (1995). Host movement and the genetic structure of populations of parasitic nematodes. *Genetics* 141 (3), 1007–1014.

Blouin, M. S., Liu, J., and Berry, R. E. (1999). Life cycle variation and the genetic structure of nematode populations. *Heredity (Edinb)* 83 (Pt 3), 253–259. doi: 10.1038/sj.hdy.6885420

Boakye, D. A., and Meredith, S. E. (1993). Introgression between members of the *Simulium damnosum* complex: larvicidal implications. *Med. Vet. Entomol.* 7 (4), 393–397. doi: 10.1111/j.1365-2915.1993.tb00714.x

Boakye, D. A. (1993). A pictorial guide to the chromosomal identification of members of the *Simulium damnosum* Theobald complex in West Africa with particular reference to the Onchocerciasis Control Programme Area. *Trop. Med. Parasitol.* 44 (3), 223–244.

Boatin, B. (2008). The Onchocerciasis Control Programme in West Africa (OCP). *Ann. Trop. Med. Parasitol.* 102 Suppl 1, 13–17. doi: 10.1179/136485908X337427

Bourguinat, C., Pion, S. D., Kamgno, J., Gardon, J., Gardon-Wendel, N., Duke, B. O., et al. (2006). Genetic polymorphism of the beta-tubulin gene of *Onchocerca volvulus* in ivermectin naive patients from Cameroon, and its relationship with fertility of the worms. *Parasitology* 132 (Pt 2), 255–262. doi: 10.1017/S0031182005008899

Bourguinat, C., Pion, S. D., Kamgno, J., Gardon, J., Duke, B. O., Boussinesq, M., et al. (2007). Genetic selection of low fertile *Onchocerca volvulus* by ivermectin treatment. *PloS Negl. Trop. Dis.* 1 (1), e72. doi: 10.1371/journal.pntd.0000072

Bourguinat, C., Ardelli, B. F., Pion, S. D., Kamgno, J., Gardon, J., Duke, B. O., et al. (2008). P-glycoprotein-like protein, a possible genetic marker for ivermectin resistance selection in *Onchocerca volvulus.* *Mol. Biochem. Parasitol.* 158 (2), 101–111. doi: 10.1016/j.molbiopara.2007.11.017

Bourguinat, C., Lee, A. C., Lizundia, R., Blagburn, B. L., Liotta, J. L., Kraus, M. S., et al. (2015). Macrocyclic lactone resistance in *Dirofilaria immitis*: failure of heartworm preventives and investigation of genetic markers for resistance. *Vet. Parasitol.* 210 (3-4), 167–178. doi: 10.1016/j.vetpar.2015.04.002

Bourguinat, C., Keller, K., Xia, J., Lepage, P., McTier, T. L., Woods, D. J., et al. (2017). Genetic profiles of ten *Dirofilaria immitis* isolates susceptible or resistant to macrocyclic lactone heartworm preventives. *Parasit. Vectors* 10 (Suppl 2), 504. doi: 10.1186/s13071-017-2428-6

Boussinesq, M., Gardon, J., Gardon-Wendel, N., and Chippaux, J. P. (2003). Clinical picture, epidemiology and outcome of *Loa*-associated serious adverse events related to mass ivermectin treatment of onchocerciasis in Cameroon. *Filaria J.* 2 (Suppl 1), S4. doi: 10.1186/1475-2883-2-S1-S4

Boussinesq, M., Fobi, G., and Kuesel, A. C. (2018). Alternative treatment strategies to accelerate the elimination of onchocerciasis. *Int. Health* 10, (suppl_1), i40–i48. doi: 10.1093/inthealth/ihx054

Cariou, M., Duret, L., and Charlat, S. (2017). The global impact of *Wolbachia* on mitochondrial diversity and evolution. *J. Evol. Biol.* 30, 2204–2210. doi: 10.1111/jeb.13186

Cheke, R. A., and Garms, R. (2013). Indices of onchocerciasis transmission by different members of the *Simulium damnosum* complex conflict with the paradigm of forest and savanna parasite strains. *Acta Trop.* 125 (1), 43–52. doi: 10.1016/j.actatropica.2012.09.002

Chesnais, C. B., Nana-Djeunga, H. C., Njamnshi, A. K., Lenou-Nanga, C. G., Boulle, C., Bissek, A. Z., et al. (2018). The temporal relationship between onchocerciasis and epilepsy: a population-based cohort study. *Lancet Infect. Dis.* 18 (11), 1278–1286. doi: 10.1016/S1473-3099(18)30425-0

Choi, Y. J., Tyagi, R., McNulty, S. N., Rosa, B. A., Ozersky, P., Martin, J., et al. (2016). Genomic diversity in *Onchocerca volvulus* and its *Wolbachia* endosymbiont. *Nat. Microbiol.* 2, 16207. doi: 10.1038/nmicrobiol.2016.207

Choi, Y. J., Bisset, S. A., Doyle, S. R., Hallsworth-Pepin, K., Martin, J., Grant, W. N., et al. (2017). Genomic introgression mapping of field-derived multiple-anthelmintic resistance in *Teladorsagia circumcincta*. *PLoS Genet.* 13 (6), e1006857. doi: 10.1371/journal.pgen.1006857

Chung, M., Small, S. T., Serre, D., Zimmerman, P. A., and Dunning Hotopp, J. C. (2017). Draft genome sequence of the *Wolbachia* endosymbiont of *Wuchereria bancrofti* wWb. *Pathog. Dis.* 75 (9), ftx115. doi: 10.1093/femspd/ftx115

Churcher, T. S., Schwab, A. E., Prichard, R. K., and Basáñez, M. G. (2008). An analysis of genetic diversity and inbreeding in *Wuchereria bancrofti*: implications for the spread and detection of drug resistance. *PLoS Negl. Trop. Dis.* 2, e211. doi: 10.1371/journal.pntd.0000211

Churcher, T. S., Pion, S. D., Osei-Atweneboana, M. Y., Prichard, R. K., Awadzi, K., Boussinesq, M., et al. (2009). Identifying sub-optimal responses to ivermectin in the treatment of River Blindness. *Proc. Natl. Acad. Sci. U.S.A.* 106 (39), 16716–16721. doi: 10.1073/pnas.0906176106

Cobo, F. (2016). Determinants of parasite drug resistance in human lymphatic filariasis. *Rev. Esp. Quimioter.* 29 (6), 288–295.

Coeli, R., Baba, E. H., Araujo, N., Coelho, P. M., and Oliveira, G. (2013). Praziquantel treatment decreases *Schistosoma mansoni* genetic diversity in experimental infections. *PLoS Negl. Trop. Dis.* 7 (12), e2596. doi: 10.1371/journal.pntd.0002596

Coffeng, L. E., Stolk, W. A., Bakker, R., and de Vlas, S. J. (2017). "Soil-transmitted helminths (STH) control, elimination, and development of drug resistance: repercussions of systematic non-participation to preventative chemotherapy," in *American Journal of Tropical Medicine and Hygiene, 2017* (Baltimore, Maryland: Amer Soc Trop Med & Hygiene).

Cotton, J. A., Bennuru, S., Grote, A., Harsha, B., Tracey, A., Beech, R., et al. (2016). The genome of *Onchocerca volvulus*, agent of river blindness. *Nat. Microbiol.* 2, 16216. doi: 10.1038/nmicrobiol.2016.216

Crawford, K. E., Hedtke, S. M., Kuesel, A. C., Doyle, S. R., Armoo, S., Osei-Atweneboana, M. Y., et al. (2019). Utility of the *Onchocerca volvulus* mitochondrial genome for delineation of parasite transmission zones. *bioRxiv*, 732446. doi: 10.1101/732446

Criscione, C. D., and Blouin, M. S. (2004). Life cycles shape parasite evolution: comparative population genetics of salmon trematodes. *Evolution* 58 (1), 198–202. doi: 10.1111/j.0014-3820.2004.tb01587.x

Criscione, C. D., Poulin, R., and Blouin, M. S. (2005). Molecular ecology of parasites: elucidating ecological and microevolutionary processes. *Mol. Ecol.* 14 (8), 2247–2257. doi: 10.1111/j.1365-294X.2005.02587.x

da Silva, E. F.Jr., de Lacerda, M. V. G., Fontes, G., Mourão, M. P. G., and Martins, M. (2017). *Wuchereria bancrofti* infection in Haitian immigrants and the risk of re-emergence of lymphatic filariasis in the Brazilian Amazon. *Rev. Soc. Bras. Med. Trop.* 50, 256–259. doi: 10.1590/0037-8682-0407-2016

Dadzie, K. Y., Remme, J., Baker, R. H., Rolland, A., and Thylefors, B. (1990). Ocular onchocerciasis and intensity of infection in the community. III. West African rainforest foci of the vector *Simulium sanctipauli*. *Trop. Med. Parasitol.* 41 (4), 376–382.

Dadzie, Y., Amazigo, U. V., Boatin, B. A., and Seketeli, A. (2018). Is onchocerciasis elimination in Africa feasible by 2025: a perspective based on lessons learnt from the African control programmes. *Infect. Dis. Poverty* 7 (1), 63. doi: 10.1186/s40249-018-0446-z

Davin-Regli, A., Abed, Y., Charrel, R. N., Bollet, C., and de Micco, P. (1995). Variations in DNA concentrations significantly affect the reproducibility of RAPD fingerprint patterns. *Res. Microbiol.* 146 (7), 561–568. doi: 10.1016/0923-2508(96)80562-6

De Sole, G., Baker, R., Dadzie, K. Y., Giese, J., Guillet, P., Keita, F. M., et al. (1991). Onchocerciasis distribution and severity in five West African countries. *Bull. World Health Organ.* 69, 689–698.

de Souza, D. K., Osei-Poku, J., Blum, J., Baidoo, H., Brown, C. A., Lawson, B. W., et al. (2014). The epidemiology of lymphatic filariasis in Ghana, explained by the possible existence of two strains of *Wuchereria bancrofti*. *Pan Afr. Med. J.* 17, 133. doi: 10.11604/pamj.2014.17.1333370

Dent, J. A., Smith, M. M., Vassilatis, D. K., and Avery, L. (2000). The genetics of ivermectin resistance in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. U.S.A.* 97 (6), 2674–2679. doi: 10.1073/pnas.97.62674

Desjardins, C. A., Cerqueira, G. C., Goldberg, J. M., Dunning Hotopp, J. C., Haas, B. J., Zucker, J., et al. (2013). Genomics of *Loa loa*, a *Wolbachia*-free filarial parasite of humans. *Nat. Genet.* 45 (5), 495–500. doi: 10.1038/ng2585

Dhamodharan, R., Das, M. K., Hoti, S. L., Das, P. K., and Dash, A. P. (2008). Genetic variability of diurnally sub-periodic *Wuchereria bancrofti* in Nicobarese tribe of Nicobar group of Islands, Andaman and Nicobar Islands, India. *Parasitol. Res.* 103 (1), 59–66. doi: 10.1007/s00436-008-0927-2

Diawara, L., Traore, M. O., Badji, A., Bissan, Y., Doumbia, K., Goita, S. F., et al. (2009). Feasibility of onchocerciasis elimination with ivermectin treatment in endemic foci in Africa: first evidence from studies in Mali and Senegal. *PLoS Negl. Trop. Dis.* 3 (7), e497. doi: 10.1371/journal.pntd.0000497

Dickson, B. F. R., Graves, P. M., and McBride, W. J. (2017). Lymphatic filariasis in mainland southeast Asia: a systematic review and meta-analysis of prevalence and disease burden. *Trop. Med. Infect. Dis.* 2 (3), 32. doi: 10.3390/tropicalmed2030

Dorkenoo, M. A., Bronzan, R., Yehadji, D., Tchalim, M., Yakpa, K., Etassoli, S., et al. (2018). Surveillance for lymphatic filariasis after stopping mass drug administration in endemic districts of Togo, 2010-2015. *Parasit. Vectors* 11 (1), 244. doi: 10.1186/s13071-018-2843-3

Doyle, S. R., and Cotton, J. A. (2019). Genome-wide approaches to investigate anthelmintic resistance. *Trends Parasitol.* 35 (4), 289–301. doi: 10.1016/j.pt.2019.01.004

Doyle, S. R., Armoo, S., Renz, A., Taylor, M. J., Osei-Atweneboana, M. Y., and Grant, W. N. (2016). Discrimination between *Onchocerca volvulus* and *O. ochengi* filarial larvae in *Simulium damnosum* (s.l.) and their distribution throughout central Ghana using a versatile high-resolution speciation assay. *Parasit. Vectors* 9 (1), 536. doi: 10.1186/s13071-016-1832-7

Doyle, S. R., Bourguinat, C., Nana-Djeunga, H. C., Kengne-Ouafo, J. A., Pion, S. D. S., Bopda, J., et al. (2017). Genome-wide analysis of ivermectin response by *Onchocerca volvulus* reveals that genetic drift and soft selective sweeps contribute to loss of drug sensitivity. *PLoS Negl. Trop. Dis.* 11 (7), e0005816. doi: 10.1371/journal.pntd.0005816

Doyle, S. R., Illingworth, C. J. R., Laing, R., Bartley, D. J., Redman, E., Martinelli, A., et al. (2019). Population genomic and evolutionary modelling analyses reveal a single major QTL for ivermectin drug resistance in the pathogenic nematode, *Haemonchus contortus*. *BMC Genomics* 20 (1), 218. doi: 10.1186/s12864-019-5592-6

Duke, B. O. (1981). Geographical aspects of onchocerciasis. *Ann. Soc. Belg. Med. Trop.* 61 (2), 179–186.

Duke, B. O. (1993). The population dynamics of *Onchocerca volvulus* in the human host. *Trop. Med. Parasitol.* 44 (2), 61–68.

Eberhard, M. L., Lowrie, R. C. Jr., and Lammie, P. J. (1988). Persistence of microfilaremia in bancroftian filariasis after diethylcarbamazine citrate therapy. *Trop. Med. Parasitol.* 39 (2), 128–130.

Eberhard, M. L., Lammie, P. J., Dickinson, C. M., and Roberts, J. M. (1991). Evidence of nonsusceptibility to diethylcarbamazine in *Wuchereria bancrofti*. *J. Infect. Dis.* 163 (5), 1157–1160. doi: 10.1093/infdis/163.51157

Eng, J. K., and Prichard, R. K. (2005). A comparison of genetic polymorphism in populations of *Onchocerca volvulus* from untreated- and ivermectin-treated patients. *Mol. Biochem. Parasitol.* 142 (2), 193–202. doi: 10.1016/j.molbiopara.2005.01.021

Eng, J. K., Blackhall, W. J., Osei-Atweneboana, M. Y., Bourguinat, C., Galazzo, D., Beech, R. N., et al. (2006). Ivermectin selection on beta-tubulin: evidence in *Onchocerca volvulus* and *Haemonchus contortus*. *Mol. Biochem. Parasitol.* 150 (2), 229–235. doi: 10.1016/j.molbiopara.2006.08.007

Erttmann, K. D., Unnasch, T. R., Greene, B. M., Albiez, E. J., Boateng, J., Denke, A. M., et al. (1987). A DNA sequence specific for forest form *Onchocerca volvulus*. *Nature* 327 (6121), 415–417. doi: 10.1038/327415a0

Erttmann, K. D., Meredith, S. E., Greene, B. M., and Unnasch, T. R. (1990). Isolation and characterization of form specific DNA sequences of *O. volvulus*. *Acta Leiden*. 59 (1-2), 253–260.

Ferrer-Admetlla, A., Liang, M., Korneliussen, T., and Nielsen, R. (2014). On detecting incomplete soft or hard selective sweeps using haplotype structure. *Mol. Biol. Evol*. 31 (5), 1275–1291. doi: 10.1093/molbev/msu077

Fischer, P., Bamuhiiga, J., Kilian, A. H., and Buttner, D. W. (1996). Strain differentiation of *Onchocerca volvulus* from Uganda using DNA probes. *Parasitology* 112 (Pt 4), 401–408. doi: 10.1017/S0031182000066634

Fobi, G., Yameogo, L., Noma, M., Aholou, Y., Koroma, J. B., Zoure, H. M., et al. (2015). Managing the fight against onchocerciasis in Africa: APOC experience. *PloS Negl. Trop. Dis*. 9 (5), e0003542. doi: 10.1371/journal.pntd.0003542

Fola, A. A., Nate, E., Abby Harrison, G. L., Barnadas, C., Hetzel, M. W., Iga, J., et al. (2018). Nationwide genetic surveillance of *Plasmodium vivax* in Papua New Guinea reveals heterogeneous transmission dynamics and routes of migration amongst subdivided populations. *Infect. Genet. Evol*. 58, 83–95. doi: 10.1016/j.meegid.2017.11.028

Foster, J., Ganatra, M., Kamal, I., Ware, J., Makarova, K., Ivanova, N., et al. (2005). The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. *PloS Biol*. 3 (4), e121. doi: 10.1371/journal.pbio.0030121

Frempong, K. K., Walker, M., Cheke, R. A., Tetevi, E. J., Gyan, E. T., Owusu, E. O., et al. (2016). Does increasing treatment frequency address suboptimal responses to ivermectin for the control and elimination of river blindness? *Clin. Infect. Dis*. 62 (11), 1338–1347. doi: 10.1093/cid/ciw144

Garms, R., Walsh, J. F., and Davies, J. B. (1979). Studies on the reinvasion of the Onchocerciasis Control Programme in the Volta River Basin by *Simulium damnosum s.l.* with emphasis on the south-western areas. *Tropenmed. Parasitol*. 30 (3), 345–362.

Gass, K., Beau de Rochars, M. V., Boakye, D., Bradley, M., Fischer, P. U., Gyapong, J., et al. (2012). A multicenter evaluation of diagnostic tools to define endpoints for programs to eliminate bancroftian filariasis. *PloS Negl. Trop. Dis*. 6 (1), e1479. doi: 10.1371/journal.pntd.0001479

Gilleard, J. S., and Redman, E. (2016). Genetic diversity and population structure of *Haemonchus contortus*. *Adv. Parasitol*. 93, 31–68. doi: 10.1016/bs.apar.2016.02.009

Grant, W. N., and Mascord, L. J. (1996). Beta-tubulin gene polymorphism and benzimidazole resistance in *Trichostrongylus colubriformis*. *Int. J. Parasitol*. 26, 71–77. doi: 10.1016/0020-7519(95)00092-5

Grant, W. N., and Whitington, G. E. (1994). Extensive DNA polymorphism within and between two strains of *Trichostrongylus colubriformis*. *Int. J. Parasitol*. 24 (5), 719–725. doi: 10.1016/0020-7519(94)90126-0

Grant, W. N. (1994). Genetic variation in parasitic nematodes and its implications. *Int. J. Parasitol*. 24 (6), 821–830. doi: 10.1016/0020-7519(94)90008-6

Grant, W. N. (2000). What is the real target for ivermectin resistance selection in *Onchocerca volvulus*? *Parasitol. Today* 16 (11), 458–459. doi: 10.1016/S0169-4758(00)01804-4 discussion 501-452.

Graves, P. M., Wood, P., and Bossin, H. C. (2016). "Lymphatic filariasis in Oceania," in *Neglected Tropical Diseases - Oceania*. Ed. A. Loukas (Switzerland: Springer, Cham), 101–142. doi: 10.1007/978-3-319-43148-2_4

Hahnel, S. R., Zdraljevic, S., Rodriguez, B. C., Zhao, Y., McGrath, P. T., and Andersen, E. C. (2018). Extreme allelic heterogeneity at a *Caenorhabditis elegans* beta-tubulin locus explains natural resistance to benzimidazoles. *PloS Pathog*. 14 (10), e1007226. doi: 10.1371/journal.ppat.1007226

Halder, J. B., Benton, J., Jule, A. M., Guerin, P. J., Olliaro, P. L., Basáñez, M. G., et al. (2017). Systematic review of studies generating individual participant data on the efficacy of drugs for treating soil-transmitted helminthiases and the case for data-sharing. *PloS Negl. Trop. Dis*. 11 (10), e0006053. doi: 10.1371/journal.pntd.0006053

Hapairai, L. K., Sang, M. A., Sinkins, S. P., and Bossin, H. C. (2013). Population studies of the filarial vector Aedes polynesiensis (Diptera: Culicidae) in two island settings of French Polynesia. *J. Med. Entomol* 50, 965–976. doi: 10.1603/ME12246

Hawdon, J. M., Li, T., Zhan, B., and Blouin, M. S. (2001). Genetic structure of populations of the human hookworm, *Necator americanus*, in China. *Mol. Ecol*. 10 (6), 1433–1437. doi: 10.1046/j.1365-294X.2001.01296.x

Hermisson, J., and Pennings, P. S. (2005). Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* 169 (4), 2335–2352. doi: 10.1534/genetics.104.036947

Higazi, T. B., Katholi, C. R., Mahmoud, B. M., Baraka, O. Z., Mukhtar, M. M., Qubati, Y. A., et al. (2001). *Onchocerca volvulus*: genetic diversity of parasite isolates from Sudan. *Exp. Parasitol*. 97 (1), 24–34. doi: 10.1006/expr.20004589

Higazi, T. B., Klion, A. D., Boussinesq, M., and Unnasch, T. R. (2004). Genetic heterogeneity in *Loa loa* parasites from southern Cameroon: a preliminary study. *Filaria J*. 3 (1), 4. doi: 10.1186/1475-2883-3-4

Hoerauf, A., Kruse, S., Brattig, N. W., Heinzmann, A., Mueller-Myhsok, B., and Deichmann, K. A. (2002). The variant Arg110Gln of human IL-13 is associated with an immunologically hyper-reactive form of onchocerciasis (sowda). *Microbes Infect*. 4, 37–42. doi: 10.1016/S1286-4579(01)01507-6

Horstick, O., and Runge-Ranzinger, S. (2018). Protection of the house against Chagas disease, dengue, leishmaniasis, and lymphatic filariasis: a systematic review. *Lancet Infect. Dis*. 18 (5), e147–e158. doi: 10.1016/S1473-3099(17)30422-X

Hoti, S. L., Thangadurai, R., Patra, K. P., and Das, P. K. (2007). Polymorphism of gp15/400 allergen gene of *W. bancrofti* from different regions of India endemic for lymphatic filariasis. *Infect. Genet. Evol*. 7 (2), 155–160. doi: 10.1016/j.meegid.2006.07.006

Hoti, S. L., Sharma, R., Mary, K. A., Dhamodharan, R., Krishnamoorthy, K., and Das, P. K. (2008a). A method for detecting microfilaraemia, filarial specific antigens and antibodies and typing of parasites for drug resistance and genotypes using finger prick blood sample. *Acta Trop*. 107, 268–271. doi: 10.1016/j.actatropica.2008.07.011

Hoti, S. L., Thangadurai, R., Dhamodharan, R., and Das, P. K. (2008b). Genetic heterogeneity of *Wuchereria bancrofti* populations at spatially hierarchical levels in Pondicherry and surrounding areas, south India. *Infect. Genet. Evol*. 8 (5), 644–652. doi: 10.1016/j.meegid.2008.06.002

Hoti, S. L., Dhamodharan, R., Subramaniyan, K., and Das, P. K. (2009). An allele specific PCR assay for screening for drug resistance among *Wuchereria bancrofti* populations in India. *Indian J. Med. Res*. 130 (2), 193–199.

Hu, M., Gasser, R. B., Abs El-Osta, Y. G., and Chilton, N. B. (2003). Structure and organization of the mitochondrial genome of the canine heartworm, *Dirofilaria immitis*. *Parasitology* 127 (Pt 1), 37–51. doi: 10.1017/S0031182003003275

Huestis, D. L., Dao, A., Diallo, M., Sanogo, Z. L., Samake, D., Yaro, A. S., et al. (2019). Windborne long-distance migration of malaria mosquitoes in the Sahel. *Nature* 574, 404–408. doi: 10.1038/s41586-019-1622-4

Ichimori, K., and Graves, P. M. (2017). Overview of PacELF—the Pacific Programme for the Elimination of Lymphatic Filariasis. *Trop. Med. Health* 45, 34. doi: 10.1186/s41182-017-0075-4

Ichimori, K., King, J. D., Engels, D., Yajima, A., Mikhailov, A., Lammie, P., et al. (2014). Global programme to eliminate lymphatic filariasis: the processes underlying programme success. *PloS Negl. Trop. Dis*. 8 (12), e3328. doi: 10.1371/journal.pntd.0003328

Irvine, M. A., Stolk, W. A., Smith, M. E., Subramanian, S., Singh, B. K., Weil, G. J., et al. (2017). Effectiveness of a triple-drug regimen for global elimination of lymphatic filariasis: a modelling study. *Lancet Infect. Dis*. 17 (4), 451–458. doi: 10.1016/S1473-3099(16)30467-4

Jan, P. L., Gracianne, C., Fournet, S., Olivier, E., Arnaud, J. F., Porte, C., et al. (2016). Temporal sampling helps unravel the genetic structure of naturally occurring populations of a phytoparasitic nematode. 1. Insights from the estimation of effective population sizes. *Evol. Appl*. 9 (3), 489–501. doi: 10.1111/eva.12352

Jarne, P., and Théron, A. (2001). Genetic structure in natural populations of flukes and snails: a practical approach and review. *Parasitology* 123 Suppl, S27–S40. doi: 10.1017/S0031182001007715

Jennison, C., Arnott, A., Tessier, N., Tavul, L., Koepfli, C., Felger, I., et al. (2015). *Plasmodium vivax* populations are more genetically diverse and less structured than sympatric *Plasmodium falciparum* populations. *PloS Negl. Trop. Dis*. 9 (4), e0003634. doi: 10.1371/journal.pntd.0003634

Jombart, T., Devillard, S., and Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet*. 11, 94. doi: 10.1186/1471-2156-11-94

Kamga, G. R., Dissak-Delon, F. N., Nana-Djeunga, H. C., Biholong, B. D., Mbigha-Ghogomu, S., Souopgui, J., et al. (2016). Still mesoendemic onchocerciasis in two Cameroonian community-directed treatment with ivermectin projects

despite more than 15 years of mass treatment. *Parasit. Vectors* 9 (1), 581. doi: 10.1186/s13071-016-1868-8

Kaplan, R. M. (2004). Drug resistance in nematodes of veterinary importance: a status report. *Trends Parasitol.* 20 (10), 477–481. doi: 10.1016/j.pt.2004.08.001

Keddie, E. M., Higazi, T., and Unnasch, T. R. (1998). The mitochondrial genome of *Onchocerca volvulus*: sequence, structure and phylogenetic analysis. *Mol. Biochem. Parasitol.* 95 (1), 111–127. doi: 10.1016/S0166-6851(98)00102-9

Keddie, E. M., Higazi, T., Boakye, D., Merriweather, A., Wooten, M. C., and Unnasch, T. R. (1999). *Onchocerca volvulus*: limited heterogeneity in the nuclear and mitochondrial genomes. *Exp. Parasitol.* 93 (4), 198–206. doi: 10.1006/expr.19994450

King, C. L., Suamani, J., Sanuku, N., Cheng, Y. C., Satofan, S., Mancuso, B., et al. (2018). A trial of a triple-drug treatment for lymphatic filariasis. *N. Engl. J. Med.* 379 (19), 1801–1810. doi: 10.1056/NEJMoa1706854

Koala, L., Nikièma, A., Post, R. J., Pare, A. B., Kafando, C. M., Drabo, F., et al. (2017). Recrudescence of onchocerciasis in the Comoé valley in Southwest Burkina Faso. *Acta Trop.* 166, 96–105. doi: 10.1016/j.actatropica.2016.11.003

Koala, L., Nikièma, A. S., Pare, A. B., Drabo, F., Toe, L. D., Belem, A. M. G., et al. (2019). Entomological assessment of the transmission following recrudescence of onchocerciasis in the Comoé Valley, Burkina Faso. *Parasit. Vectors* 12 (1), 34. doi: 10.1186/s13071-019-3290-5

Komlan, K., Vossberg, P. S., Gantin, R. G., Solim, T., Korbmacher, F., Banla, M., et al. (2018). *Onchocerca volvulus* infection and serological prevalence, ocular onchocerciasis and parasite transmission in northern and central Togo after decades of *Simulium damnosum s.l.* vector control and mass drug administration of ivermectin. *PloS Negl. Trop. Dis.* 12 (3), e0006312. doi: 10.1371/journal.pntd.0006312

Krueger, A., and Hennings, I. C. (2006). Molecular phylogenetics of blackflies of the *Simulium damnosum* complex and cytophylogenetic implications. *Mol. Phylogenet. Evol.* 39 (1), 83–90. doi: 10.1016/j.ympev.2005.11.007

Kumar, N. P., Patra, K. P., Hoti, S. L., and Das, P. K. (2002). Genetic variability of the human filarial parasite, *Wuchereria bancrofti* in South India. *Acta Trop.* 82, 67–76. doi: 10.1016/S0001-706X(02)00037-2

Kumar, A., Mhatre, S., Godbole, S., Jha, P., and Dikshit, R. (2019). Optimization of extraction of genomic DNA from archived dried blood spot (DBS): potential application in epidemiological research & bio banking. *Gates Open Res.* 2, 57. doi: 10.12688/gatesopenres.12855.2

Kwa, M. S., Veenstra, J. G., and Roos, M. H. (1994). Benzimidazole resistance in *Haemonchus contortus* is correlated with a conserved mutation at amino acid 200 in beta-tubulin isotype 1. *Mol. Biochem. Parasitol.* 63 (2), 299–303. doi: 10.1016/0166-6851(94)90066-3

Lau, C. L., Won, K. Y., Becker, L., Soares Magalhaes, R. J., Fuimaono, S., Melrose, W., et al. (2014). Seroprevalence and spatial epidemiology of lymphatic filariasis in American Samoa after successful mass drug administration. *PloS Negl. Trop. Dis.* 8 (11), e3297. doi: 10.1371/journal.pntd.0003297

Lau, C. L., Won, K. Y., Lammie, P. J., and Graves, P. M. (2016). Lymphatic filariasis elimination in American Samoa: evaluation of molecular xenomonitoring as a surveillance tool in the endgame. *PloS Negl. Trop. Dis.* 10 (11), e0005108. doi: 10.1371/journal.pntd.0005108

Lau, C. L., Sheridan, S., Ryan, S., Roineau, M., Andreosso, A., Fuimaono, S., et al. (2017). Detecting and confirming residual hotspots of lymphatic filariasis transmission in American Samoa 8 years after stopping mass drug administration. *PloS Negl. Trop. Dis.* 11 (9), e0005914. doi: 10.1371/journal.pntd.0005914

Laval, G., and Excoffier, L. (2004). SIMCOAL 2.0 : a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. *Bioinformatics* 20, 2485–2487. doi: 10.1093/bioinformatics/bth264

Le Berre, R., Garms, R., Davies, J. B., Walsh, J. F., and Philippon, B. (1979). Displacements of *Simulium damnosum* and strategy of control of control against onchocerciasis. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 287, 277–288. doi: 10.1098/rstb.19790061

Levins, R. (1969). Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bull. Entomol Soc. Am.* 15 (3), 237–240. doi: 10.1093/besa/15.3.237

Lo, E., Lam, N., Hemming-Schroeder, E., Nguyen, J., Zhou, G., Lee, M. C., et al. (2017). Frequent spread of *Plasmodium vivax* malaria maintains high genetic diversity at the Myanmar-China Border, without distance and landscape barriers. *J. Infect. Dis.* 216 (10), 1254–1263. doi: 10.1093/infdis/jix106

Lok, J. B., Knight, D. H., Selavka, C. M., Eynard, J., Zhang, Y., and Bergman, R. N. (1995). Studies of reproductive competence in male *Dirofilaria immitis* treated with milbemycin oxime. *Trop. Med. Parasitol.* 46 (4), 235–240.

Lont, Y. L., Coffeng, L. E., de Vlas, S. J., Golden, A., de Los Santos, T., Domingo, G. J., et al. (2017). Modelling anti-Ov16 IgG4 antibody prevalence as an indicator for evaluation and decision making in onchocerciasis elimination programmes. *PloS Negl. Trop. Dis.* 11 (1), e0005314. doi: 10.1371/journal.pntd.0005314

Mahakalkar, A. L., Sapkal, H. P., and Baig, M. M. (2017). Report of high genetic diversity of filarial worm, *Wuchereria bancrofti* from endemic region of Eastern Maharashtra (India). *Helminthologia* 54, 292–297. doi: 10.1515/helm-2017-0043

McCarthy, J. (2005). Is anthelmintic resistance a threat to the program to eliminate lymphatic filariasis? *Am. J. Trop. Med. Hyg.* 73 (2), 232–233. doi: 10.4269/ajtmh.2005.73.232

McCulloch, K., McCaw, J., McVernon, J., Hedtke, S. M., Walker, M., Milton, P., et al. (2017). "Investigation into the effect of host migration on the transmission of *Onchocerca volvulus* using a patch model," in *American Journal of Tropical Medicine and Hygiene, 2017* (Baltimore, Maryland: Amer Soc Trop Med & Hygiene).

McNulty, S. N., Mullin, A. S., Vaughan, J. A., Tkach, V. V., Weil, G. J., and Fischer, P. U. (2012). Comparing the mitochondrial genomes of *Wolbachia*-dependent and independent filarial nematode species. *BMC Genomics* 13, 145. doi: 10.1186/1471-2164-13-145

Menez, C., Alberich, M., Kansoh, D., Blanchard, A., and Lespine, A. (2016). Acquired tolerance to ivermectin and moxidectin after drug selection pressure in the nematode *Caenorhabditis elegans*. *Antimicrob. Agents Chemother.* 60 (8), 4809–4819. doi: 10.1128/AAC.00713-16

Morales-Hojas, R., Cheke, R. A., and Post, R. J. (2007). A preliminary analysis of the population genetics and molecular phylogenetics of *Onchocerca volvulus* (Nematoda: Filarioidea) using nuclear ribosomal second internal transcribed spacer sequences. *Mem. Inst. Oswaldo Cruz* 102 (7), 879–882. doi: 10.1590/S0074-02762007005000114

Mukhtar, M. M., Khier, M. M., Baraka, O. Z., and Homeida, M. M. (1998). The burden of *Onchocerca volvulus* in Sudan. *Ann. Trop. Med. Parasitol.* 92 (Suppl 1), S129–S131. doi: 10.1080/00034983.1998.11813376

Nadler, S. A. (1995). Microevolution and the genetic structure of parasite populations. *J. Parasitol.* 81 (3), 395–403. doi: 10.2307/3283821

Nana-Djeunga, H. C., Bourguinat, C., Pion, S. D., Bopda, J., Kengne-Ouafo, J. A., Njiokou, F., et al. (2014). Reproductive status of *Onchocerca volvulus* after ivermectin treatment in an ivermectin-naive and a frequently treated population from Cameroon. *PloS Negl. Trop. Dis.* 8 (4), e2824. doi: 10.1371/journal.pntd.0002824

Nana-Djeunga, H. C., Fossuo-Thotchum, F., Pion, S. D., Chesnais, C. B., Kubofcik, J., Mackenzie, C. D., et al. (2019). *Loa loa* microfilariae in skin snips: consequences for onchocerciasis monitoring and evaluation in *L. loa*-endemic areas. *Clin. Infect. Dis.* 69 (9), 1628–1630. doi: 10.1093/cid/ciz172

Nikièma, A. S., Koala, L., Post, R. J., Pare, A. B., Kafando, C. M., Drabo, F., et al. (2018). Onchocerciasis prevalence, human migration and risks for onchocerciasis elimination in the Upper Mouhoun, Nakambé and Nazinon river basins in Burkina Faso. *Acta Trop.* 185, 176–182. doi: 10.1016/j.actatropica.2018.05.013

Noma, M., Nwoke, B. E., Nutall, I., Tambala, P. A., Enyong, P., Namsenmo, A., et al. (2002). Rapid epidemiological mapping of onchocerciasis (REMO): its application by the African Programme for Onchocerciasis Control (APOC). *Ann. Trop. Med. Parasitol.* 96 (Suppl 1), S29–S39. doi: 10.1179/000349802125000637

Noma, M., Zoure, H. G., Tekle, A. H., Enyong, P. A., Nwoke, B. E., and Remme, J. H. (2014). The geographic distribution of onchocerciasis in the 20 participating countries of the African Programme for Onchocerciasis Control: (1) priority areas for ivermectin treatment. *Parasit. Vectors* 7, 325. doi: 10.1186/1756-3305-7-325

Nuchprayoon, S., Junpee, A., and Poovorawan, Y. (2007). Random amplified polymorphic DNA (RAPD) for differentiation between Thai and Myanmar strains of *Wuchereria bancrofti*. *Filaria J.* 6, 6. doi: 10.1186/1475-2883-6-6

Nunes, L. V., Rocha, A., Araújo, J., Braga, C., Alcantar, P., Fiorillo, K., et al. (2016). Lymphatic filariasis: surveillance action among immigrants from endemic areas, Acre State, Brazilian Amazon. *Asian Pac. J. Trop. Dis.* 6, 521–526. doi: 10.1016/S2222-1808(16)61081-2

Ogunrinade, A., Boakye, D., Merriweather, A., and Unnasch, T. R. (1999). Distribution of the blinding and nonblinding strains of *Onchocerca volvulus* in Nigeria. *J. Infect. Dis.* 179 (6), 1577–1579. doi: 10.1086/314784

Opoku, N. O., Bakajika, D. K., Kanza, E. M., Howard, H., Mambandu, G. L., Nyathirombo, A., et al. (2018). Single dose moxidectin versus ivermectin for *Onchocerca volvulus* infection in Ghana, Liberia, and the Democratic Republic of the Congo: a randomised, controlled, double-blind phase 3 trial. *Lancet* 392 (10154), 1207–1216. doi: 10.1016/S0140-6736(17)32844-1

Osei-Atweneboana, M. Y., Eng, J. K., Boakye, D. A., Gyapong, J. O., and Prichard, R. K. (2007). Prevalence and intensity of *Onchocerca volvulus* infection and efficacy of ivermectin in endemic communities in Ghana: a two-phase epidemiological study. *Lancet* 369 (9578), 2021–2029. doi: 10.1016/S0140-6736(07)60942-8

Osei-Atweneboana, M. Y., Awadzi, K., Attah, S. K., Boakye, D. A., Gyapong, J. O., and Prichard, R. K. (2011). Phenotypic evidence of emerging ivermectin resistance in *Onchocerca volvulus*. *PloS Negl. Trop. Dis.* 5 (3), e998. doi: 10.1371/journal.pntd.0000998

Osei-Atweneboana, M. Y., Boakye, D. A., Awadzi, K., Gyapong, J. O., and Prichard, R. K. (2012). Genotypic analysis of beta-tubulin in *Onchocerca volvulus* from communities and individuals showing poor parasitological response to ivermectin treatment. *Int. J. Parasitol. Drugs Drug Resist.* 2, 20–28. doi: 10.1016/j.ijpddr.2012.01.005

Owusu, I. O., de Souza, D. K., Anto, F., Wilson, M. D., Boakye, D. A., Bockarie, M. J., et al. (2015). Evaluation of human and mosquito based diagnostic tools for defining endpoints for elimination of *Anopheles* transmitted lymphatic filariasis in Ghana. *Trans. R Soc. Trop. Med. Hyg.* 109 (10), 628–635. doi: 10.1093/trstmh/trv070

Pérez, T., Albornoz, J., and Dominguez, A. (1998). An evaluation of RAPD fragment reproducibility and nature. *Mol. Ecol.* 7 (10), 1347–1357. doi: 10.1046/j.1365-294x.1998.00484.x

Patra, K. P., Ramu, T., Hoti, S. L., Pragasam, G. S., and Das, P. K. (2007). Identification of a molecular marker for genotyping human lymphatic filarial nematode parasite *Wuchereria bancrofti*. *Exp. Parasitol.* 116 (1), 59–65. doi: 10.1016/j.exppara.2006.11.011

Peng, B., and Kimmel, M. (2005). simuPOP : a forward-time population genetics simulation environment. *Bioinformatics* 21, 3686–3687. doi: 10.1093/bioinformatics/bti584

Penner, G. A., Bush, A., Wise, R., Kim, W., Domier, L., Kasha, K., et al. (1993). Reproducibility of random amplified polymorphic DNA (RAPD) analysis among laboratories. *PCR Methods Appl.* 2 (4), 341–345. doi: 10.1101/gr.2.4.341

Pennings, P. S., and Hermisson, J. (2006). Soft sweeps III: the signature of positive selection from recurrent mutation. *PloS Genet.* 2 (12), e186. doi: 10.1371/journal.pgen.0020186

Pichon, G. (2002). Limitation and facilitation in the vectors and other aspects of the dynamics of filarial transmission: the need for vector control against *Anopheles*-transmitted filariasis. *Ann. Trop. Med. Parasitol.* 96 (Suppl 2), S143–A152. doi: 10.1179/000349802125002509

Pion, S. D., Nana-Djeunga, H. C., Kamgno, J., Tendongfor, N., Wanji, S., Njiokou, F., et al. (2013). Dynamics of *Onchocerca volvulus* microfilarial densities after ivermectin treatment in an ivermectin-naive and a multiply treated population from Cameroon. *PloS Negl. Trop. Dis.* 7 (2), e2084. doi: 10.1371/journal.pntd.0002084

Pion, S. D., Montavon, C., Chesnais, C. B., Kamno, J., Wanji, S., Klion, A. D., et al. (2016). Positivity of antigen tests used for diagnosis of lymphatic filariasis *Wuchereria bancrofti* infection but with high *Loa loa* microfilaremia. *Am. J. Trop. Med. Hyg.* 95 (6), 1417–1423. doi: 10.4269/ajtmh.16-0547

Prichard, R. (2001). Genetic variability following selection of *Haemonchus contortus* with anthelmintics. *Trends Parasitol.* 17 (9), 445–453. doi: 10.1016/S1471-4922(01)01983-3

Pritchard, J. K., and Di Rienzo, A. (2010). Adaptation - not by sweeps alone. *Nat. Rev. Genet.* 11 (10), 665–667. doi: 10.1038/nrg2880

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.

Prost, A., Hervouet, J. P., and Thylefors, B. (1979). Epidemiologic status of onchocerciasis. *Bull. World Health Organ.* 57 (4), 655–662.

Prugnolle, F., Liu, H., de Meeus, T., and Balloux, F. (2005). Population genetics of complex life-cycle parasites: an illustration with trematodes. *Int. J. Parasitol.* 35 (3), 255–263. doi: 10.1016/j.ijpara.2004.10.027

Przeworski, M., Coop, G., and Wall, J. D. (2005). The signature of positive selection on standing genetic variation. *Evolution* 59 (11), 2312–2323. doi: 10.1554/05-273.1

Ramaiah, K. D., and Ottesen, E. A. (2014). Progress and impact of 13 years of the global programme to eliminate lymphatic filariasis on reducing the burden of filarial disease. *PloS Negl. Trop. Dis.* 8 (11), e3319. doi: 10.1371/journal.pntd.0003319

Ramaiah, K. D. (2013). Population migration: implications for lymphatic filariasis elimination programmes. *PloS Negl. Trop. Dis.* 7 (3), e2079. doi: 10.1371/journal.pntd.0002079

Ramesh, A., Small, S. T., Kloos, Z. A., Kazura, J. W., Nutman, T. B., Serre, D., et al. (2012). The complete mitochondrial genome sequence of the filarial nematode *Wuchereria bancrofti* from three geographic isolates provides evidence of complex demographic history. *Mol. Biochem. Parasitol.* 183 (1), 32–41. doi: 10.1016/j.molbiopara.2012.01.004

Rao, R. U., Atkinson, L. J., Ramzy, R. M., Helmy, H., Farid, H. A., Bockarie, M. J., et al. (2006). A real-time PCR-based assay for detection of *Wuchereria bancrofti* DNA in blood and mosquitoes. *Am. J. Trop. Med. Hyg.* 74 (5), 826–832. doi: 10.4269/ajtmh.2006.74.826

Rao, R. U., Nagodavithana, K. C., Samarasekera, S. D., Wijegunawardana, A. D., Premakumara, W. D., Perera, S. N., et al. (2014). A comprehensive assessment of lymphatic filariasis in Sri Lanka six years after cessation of mass drug administration. *PloS Negl. Trop. Dis.* 8 (11), e3281. doi: 10.1371/journal.pntd.0003281

Rao, R. U., Samarasekera, S. D., Nagodavithana, K. C., Dassanayaka, T. D. M., Punchihewa, M. W., Ranasinghe, U. S. B., et al. (2017). Reassessment of areas with persistent lymphatic filariasis nine years after cessation of mass drug administration in Sri Lanka. *PloS Negl. Trop. Dis.* 11 (10), e0006066. doi: 10.1371/journal.pntd.0006066

Rao, R. U., Samarasekera, S. D., Nagodavithana, K. C., Goss, C. W., Punchihewa, M. W., Dassanayaka, T. D. M., et al. (2018). Comprehensive assessment of a hotspot with persistent Bancroftian filariasis in coastal Sri Lanka. *Am. J. Trop. Med. Hyg.* 99 (3), 735–742. doi: 10.4269/ajtmh.18-0169

Rebollo, M. P., Zoure, H., Ogoussan, K., Sodahlon, Y., Ottesen, E. A., and Cantey, P. T. (2018). Onchocerciasis: shifting the target from control to elimination requires a new first-step elimination mapping. *Int. Health* 10 (suppl_1), i14–i19. doi: 10.1093/inthealth/ihx052

Redman, E., Whitelaw, F., Tait, A., Burgess, C., Bartley, Y., Skyce, P. J., et al. (2015). The emergence of resistance to the benzimidazole anthelmintics in parasitic nematodes of livestock is characterised by multiple independent hard and soft selective sweeps. *PloS Negl. Trop. Dis.* 9 (2), e0003494. doi: 10.1371/journal.pntd.0003494

Remme, J., Dadzie, K. Y., Rolland, A., and Thylefors, B. (1989). Ocular onchocerciasis and intensity of infection in the community. I. West African savanna. *Trop. Med. Parasitol.* 40 (3), 340–347.

Remme, J. H. F., Alley, E. S., and Plaisier, A. P. (1995). "Estimation and prediction in tropical disease control: the example of onchocerciasis," in *Epidemic models: their structure and relation to data*. Ed. D. Mollison (Cambridge: Cambridge University Press).

Remme, J. H. F., Boatin, B., and Boussinesq, M. (2017). "Helminthic diseases: onchocerciasis and loiasis," in *The International Encyclopedia of Public Health, 2 Edn.* Eds. S. R. Quah and W. C. Cockerham (Amsterdam: Elsevier). doi: 10.1016/B978-0-12-803678-5.00206-X

Remme, J. H. (2004). Research for control: the onchocerciasis experience. *Trop. Med. Int. Health* 9 (2), 243–254. doi: 10.1046/j.1365-3156.2003.01192.x

Rodelsperger, C., Neher, R. A., Weller, A. M., Eberhardt, G., Witte, H., Mayer, W. E., et al. (2014). Characterization of genetic diversity in the nematode *Pristionchus pacificus* from population-scale resequencing data. *Genetics* 196 (4), 1153–1165. doi: 10.1534/genetics.113.159855

Rodriguez-Pérez, M. A., Katholi, C. R., Hassan, H. K., and Unnasch, T. R. (2006). Large-scale entomologic assessment of *Onchocerca volvulus* transmission by poolscreen PCR in Mexico. *Am. J. Trop. Med. Hyg.* 74 (6), 1026–1033. doi: 10.4269/ajtmh.2006.741026

Satimai, W., Jiraamonnimit, C., Thammapalo, S., Choochote, W., Luenee, P., Boitano, J. J., et al. (2011). The impact of a national program to eliminate lymphatic filariasis in selected Myanmar immigrant communities in Bangkok and Ranong Province, Thailand. *Southeast Asian J. Trop. Med. Public Health* 42 (5), 1054–1064.

Sauerbrey, M., Rakers, L. J., and Richards, F. O. (2018). Progress toward elimination of onchocerciasis in the Americas. *Int. Health* 10 ((suppl_1)), i71–i78. doi: 10.1093/inthealth/ihx039

Schwab, A. E., Boakye, D. A., Kyelem, D., and Prichard, R. K. (2005). Detection of benzimidazole resistance-associated mutations in the filarial nematode *Wuchereria bancrofti* and evidence for selection by albendazole and ivermectin combination treatment. *Am. J. Trop. Med. Hyg.* 73 (2), 234–238. doi: 10.4269/ajtmh.2005.73.234

Schwab, A. E., Churcher, T. S., Schwab, A. J., Basáñez, M. G., and Prichard, R. K. (2006). Population genetics of concurrent selection with albendazole and ivermectin or diethylcarbamazine on the possible spread of albendazole resistance in *Wuchereria bancrofti*. *Parasitology* 133 (Pt 5), 589–601. doi: 10.1017/S003118200600076X

Schwab, A. E., Churcher, T. S., Schwab, A. J., Basáñez, M. G., and Prichard, R. K. (2007). An analysis of the population genetics of potential multi-drug resistance in *Wuchereria bancrofti* due to combination chemotherapy. *Parasitology* 134 (Pt 7), 1025–1040. doi: 10.1017/S0031182007002363

Schwabl, P., Llewellyn, S., Landguth, E. L., Andersson, B., Kitron, U., Costales, J. A., et al. (2017). Prediction and prevention of parasitic diseases using a landscape genomics framework. *Trends Parasitol.* 33 (4), 264–275. doi: 10.1016/j.pt.2016.10.008

Seketeli, A., Adeoye, G., Eyamba, A., Nnoruka, E., Drameh, P., Amazigo, U. V., et al. (2002). The achievements and challenges of the African Programme for Onchocerciasis Control (APOC). *Ann. Trop. Med. Parasitol.* 96 (Suppl 1), S15–S28. doi: 10.1179/000349802125000628

Seketeli, A. (2002). APOC at mid-point: so far so good. *Ann. Trop. Med. Parasitol.* 96 (Suppl 1), S3–S4. doi: 10.1179/atm.2002.96.supplement1.001

Silvestre, A., and Cabaret, J. (2002). Mutation in position 167 of isotype 1 beta-tubulin gene of Trichostrongylid nematodes: role in benzimidazole resistance? *Mol. Biochem. Parasitol.* 120 (2), 297–300. doi: 10.1016/S0166-6851(01)00455-8

Silvestre, A., and Humbert, J. F. (2002). Diversity of benzimidazole-resistance alleles in populations of small ruminant parasites. *Int. J. Parasitol.* 32 (7), 921–928. doi: 10.1016/S0020-7519(02)00032-2

Small, S. T., Ramesh, A., Bun, K., Reimer, L., Thomsen, E., Baea, M., et al. (2013). Population genetics of the filarial worm *Wuchereria bancrofti* in a post-treatment region of Papua New Guinea: insights into diversity and life history. *PloS Negl. Trop. Dis.* 7 (7), e2308. doi: 10.1371/journal.pntd.0002308

Small, S. T., Tisch, D. J., and Zimmerman, P. A. (2014). Molecular epidemiology, phylogeny and evolution of the filarial nematode *Wuchereria bancrofti*. *Infect. Genet. Evol.* 28, 33–43. doi: 10.1016/j.meegid.2014.08.018

Small, S. T., Reimer, L. J., Tisch, D. J., King, C. L., Christensen, B. M., Siba, P. M., et al. (2016). Population genomics of the filarial nematode parasite *Wuchereria bancrofti* from mosquitoes. *Mol. Ecol.* 25 (7), 1465–1477. doi: 10.1111/mec.13574

Small, S. T., Labbé, F., Coulibaly, Y. I., Nutman, T. B., King, C. L., Serre, D., et al. (2019). Human migration and the spread of the nematode parasite *Wuchereria bancrofti*. *Mol. Biol. Evol.* 36 (9), 1931–1941. doi: 10.1093/molbev/msz116

Smith, M. E., Singh, B. K., Irvine, M. A., Stolk, W. A., Subramanian, S., Hollingsworth, T. D., et al. (2017). Predicting lymphatic filariasis transmission and elimination dynamics using a multi-model ensemble framework. *Epidemics* 18, 16–28. doi: 10.1016/j.epidem.2017.02.006

Stolk, W. A., Walker, M., Coffeng, L. E., Basáñez, M. G., and de Vlas, S. J. (2015). Required duration of mass ivermectin treatment for onchocerciasis elimination in Africa: a comparative modelling analysis. *Parasit. Vectors* 8, 552. doi: 10.1186/s13071-015-1159-9

Tadiri, C. P., Scott, M. E., and Fussmann, G. F. (2018). Microparasite dispersal in metapopulations: a boon or bane to the host population? *Proc. Biol. Sci.* 29, 285. doi: 10.1098/rspb.20181519

Tang, J., Toe, L., Back, C., Zimmerman, P. A., Pruess, K., and Unnasch, T. R. (1995). The *Simulium damnosum* species complex: phylogenetic analysis and molecular identification based upon mitochondrially encoded gene sequences. *Insect Mol. Biol.* 4 (2), 79–88. doi: 10.1111/j.1365-2583.1995.tb00011.x

Tang, J., Toe, L., Back, C., and Unnasch, T. R. (1996). Intra-specific heterogeneity of the rDNA internal transcribed spacer in the *Simulium damnosum* (Diptera: Simuliidae) complex. *Mol. Biol. Evol.* 13 (1), 244–252. doi: 10.1093/oxfordjournals.molbev.a025561

Taylor, M. J., Hoerauf, A., and Bockarie, M. (2010). Lymphatic filariasis and onchocerciasis. *Lancet* 376 (9747), 1175–1185. doi: 10.1016/S0140-6736(10)60586-7

Tekle, A. H., Zoure, H. G., Noma, M., Boussinesq, M., Coffeng, L. E., Stolk, W. A., et al. (2016). Progress towards onchocerciasis elimination in the participating countries of the African Programme for Onchocerciasis Control: epidemiological evaluation results. *Infect. Dis. Poverty* 5 (1), 66. doi: 10.1186/s40249-016-0160-7

Thangadurai, R., Hoti, S. L., Kumar, N. P., and Das, P. K. (2006). Phylogeography of human lymphatic filarial parasite, *Wuchereria bancrofti* in India. *Acta Trop.* 98 (3), 297–304. doi: 10.1016/j.actatropica.2006.06.004

Toothong, T., Tipayamongkholgul, M., Suwannapong, N., and Suvannadabba, S. (2015). Evaluation of mass drug administration in the program to control imported lymphatic filariasis in Thailand. *BMC Public Health* 15, 975. doi: 10.1186/s12889-015-2325-x

Traore, M. O., Sarr, M. D., Badji, A., Bissan, Y., Diawara, L., Doumbia, K., et al. (2012). Proof-of-principle of onchocerciasis elimination with ivermectin treatment in endemic foci in Africa: final results of a study in Mali and Senegal. *PloS Negl. Trop. Dis.* 6 (9), e1825. doi: 10.1371/journal.pntd.0001825

Turner, H. C., Churcher, T. S., Walker, M., Osei-Atweneboana, M. Y., Prichard, R. K., and Basáñez, M. G. (2013). Uncertainty surrounding projections of the long-term impact of ivermectin treatment on human onchocerciasis. *PloS Negl. Trop. Dis.* 7 (4), e2169. doi: 10.1371/journal.pntd.0002169

UNDP/World Bank/WHO Special Programme for Research & Training in Tropical Diseases. (1997). *Tropical disease research: progress 1995-96: thirteenth programme report* (Geneva: World Health Organization).

UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases. (1992). *Methods for community diagnosis of onchocerciasis to guide ivermectin based control in Africa: report of an informal consultation held in Ouagadougou from 19-21 November 1991* (Geneva: World Health Organization).

Unnasch, T. R., and Williams, S. A. (2000). The genomes of *Onchocerca volvulus*. *Int. J. Parasitol.* 30 (4), 543–552. doi: 10.1016/S0020-7519(99)00184-8

Verdonschot, P. F. M., and Besse-Lototskaya, A. A. (2014). Flight distance of mosquitoes (Culicidae): a metadata analysis to support the management of barrier zones around rewetted and newly constructed wetlands. *Limnologica* 45, 69–79. doi: 10.1016/j.limno.2013.11.002

Verver, S., Walker, M., Kim, Y. E., Fobi, G., Tekle, A. H., Zoure, H. G. M., et al. (2018). How can onchocerciasis elimination in Africa be accelerated? Modeling the impact of increased ivermectin treatment frequency and complementary vector control. *Clin. Infect. Dis.* 66 (suppl_4), S267–S274. doi: 10.1093/cid/cix1137

Walker, M., Mabud, T. S., Olliaro, P. L., Coulibaly, J. T., King, C. H., Raso, G., et al. (2016). New approaches to measuring anthelminthic drug efficacy: parasitological responses of childhood schistosome infections to treatment with praziquantel. *Parasit. Vectors* 9, 41. doi: 10.1186/s13071-016-1312-0

Walker, M., Stolk, W. A., Dixon, M. A., Bottomley, C., Diawara, L., Traore, M. O., et al. (2017). Modelling the elimination of river blindness using long-term epidemiological and programmatic data from Mali and Senegal. *Epidemics* 18, 4–15. doi: 10.1016/j.epidem.2017.02.005

Waltmann, A., Koepfli, C., Tessier, N., Karl, S., Fola, A., Darcy, A. W., et al. (2018). Increasingly inbred and fragmented populations of *Plasmodium vivax* associated with the eastward decline in malaria transmission across the Southwest Pacific. *PloS Negl. Trop. Dis.* 12 (1), e0006146. doi: 10.1371/journal.pntd.0006146

Wanji, S., Amvongo-Adjia, N., Koudou, B., Njouendou, A. J., Chounna Ndongmo, P. W., Kengne-Ouafo, J. A., et al. (2015). Cross-reactivity of filariasis ICT cards in areas of contrasting endemicity of *Loa loa* and *Mansonella perstans* in Cameroon: implications for shrinking of the lymphatic filariasis map in the Central African Region. *PloS Negl. Trop. Dis.* 9 (11), e004184. doi: 10.1371/journal.pntd.0004184

Weigand, H., and Leese, F. (2018). Detecting signatures of positive selection in non-model species using genomic data. *Zool. J. Linn. Soc.* 184, 528–583. doi: 10.1093/zoolinnean/zly007

Wesolowski, A., Taylor, A. R., Chang, H.-H., Verity, R., Tessema, S., Bailey, J. A., et al. (2018). Mapping malaria by combining parasite genomic and epidemiologic data. *BMC Med.* 16, 190. doi: 10.1186/s12916-018-1232-2

WHO Expert Committee on Onchocerciasis Control (1993) and World Health Organization (1995). "Onchocerciasis and its control: report of a WHO Expert Committee on Onchocerciasis Control", in: *WHO technical report series* (Geneva: World Health Organization).

Williams, S. A., Lizotte-Waniewski, M. R., Foster, J., Guiliano, D., Daub, J., Scott, A. L., et al. (2000). The filarial genome project: analysis of the nuclear, mitochondrial and endosymbiont genomes of *Brugia malayi*. *Int. J. Parasitol.* 30 (4), 411–419. doi: 10.1016/S0020-7519(00)00014-X

Wolstenholme, A. J., Fairweather, I., Prichard, R., von Samson-Himmelstjerna, G., and Sangster, N. C. (2004). Drug resistance in veterinary helminths. *Trends Parasitol.* 20 (10), 469–476. doi: 10.1016/j.pt.2004.07.010

Wolstenholme, A. J., Evans, C. C., Jimenez, P. D., and Moorhead, A. R. (2015). The emergence of macrocyclic lactone resistance in the canine heartworm, *Dirofilaria immitis*. *Parasitology* 142 (10), 1249–1259. doi: 10.1017/S003118201500061X

World Health Organization. (2001). *Certification of elimination of human onchocerciasis: criteria and procedures* (Geneva: World Health Organization).

World Health Organization. (2011). *Monitoring and epidemiological assessment of mass drug administration: a manual for national elimination programmes* (Geneva: World Health Organization).

World Health Organization. (2012a). *Accelerating work to overcome the global impact of neglected tropical diseases: a roadmap for implementation: executive summary* (Geneva: World Health Organization). https://apps.who.int/iris/handle/10665/70809

World Health Organization. (2012b). *Provisional strategy for interrupting lymphatic filariasis transmission in loiasis-endemic countries: report of the meeting on lymphatic filariasis, malaria and integrated vector management, Accra, Ghana, 5–9 March 2012*. Ed. K. Ichimori (Geneva: World Health Organization).

World Health Organization. (2013). *Lymphatic filariasis: a handbook of practical entomology for national lymphatic filariasis elimination programmes* (Geneva: World Health Organization).

World Health Organization. (2016). *Guidelines for stopping mass drug administration and verifying elimination of human onchocerciasis: criteria and procedures* (Geneva: World Health Organization).

World Health Organization. (2017a). *Guideline - Alternative mass drug administration regimens to eliminate lymphatic filariasis* (Geneva: World Health Organization).

World Health Organization. (2017b). *Validation of elimination of lymphatic filariasis as a public health problem* (Geneva: World Health Organization).

World Health Organization. (2018). Progress report on the elimination of human onchocerciasis, 2017-2018. *Wkly. Epidemiol. Rec.* 93, 633–643.

World Health Organization. (2019). Global Program to eliminate lymphatic filariasis: progress report, 2018. *Wkly. Epidemiol. Rec.* 94, 457–469.

Xu, Z., Graves, P. M., Lau, C. L., Clements, A., Geard, N., and Glass, K. (2018a). GEOFIL: A spatially-explicit agent-based modelling framework for predicting the long-term transmission dynamics of lymphatic filariasis in American Samoa. *Epidemics* 27, 19–27. doi: 10.1016/j.epidem.2018.12.003

Xu, Z., Lau, C. L., Zhou, X., Fuimaono, S., Soares Magalhaes, R. J., and Graves, P. M. (2018b). The extensive networks of frequent population mobility in the Samoan Islands and their implications for infectious disease transmission. *Sci. Rep.* 8 (1), 10136. doi: 10.1038/s41598-018-28081-x

Yan, R., Urdaneta-Marquez, L., Keller, K., James, C. E., Davey, M. W., and Prichard, R. K. (2012). The role of several ABC transporter genes in ivermectin resistance in *Caenorhabditis elegans*. *Vet. Parasitol.* 190 (3-4), 519–529. doi: 10.1016/j.vetpar.2012.06.038

Zhou, X., Xie, Y., Zhang, Z. H., Wang, C. D., Sun, Y., Gu, X. B., et al. (2013). Analysis of the genetic diversity of the nematode parasite *Baylisascaris schroederi* from wild giant pandas in different mountain ranges in China. *Parasit. Vectors* 6, 233. doi: 10.1186/1756-3305-6-233

Zimmerman, P. A., Dadzie, K. Y., De Sole, G., Remme, J., Alley, E. S., and Unnasch, T. R. (1992). *Onchocerca volvulus* DNA probe classification correlates with epidemiologic patterns of blindness. *J. Infect. Dis.* 165 (5), 964–968. doi: 10.1093/infdis/165.5.964

Zimmerman, P. A., Toe, L., and Unnasch, T. R. (1993). Design of *Onchocerca* DNA probes based upon analysis of a repeated sequence family. *Mol. Biochem. Parasitol.* 58 (2), 259–267. doi: 10.1016/0166-6851(93)90047-2

Zimmerman, P. A., Guderian, R. H., Aruajo, E., Elson, L., Phadke, P., Kubofcik, J., et al. (1994a). Polymerase chain reaction-based diagnosis of *Onchocerca volvulus* infection: improved detection of patients with onchocerciasis. *J. Infect. Dis.* 169 (3), 686–689. doi: 10.1093/infdis/169.3.686

Zimmerman, P. A., Katholi, C. R., Wooten, M. C., Lang-Unnasch, N., and Unnasch, T. R. (1994b). Recent evolutionary history of American *Onchocerca volvulus*, based on analysis of a tandemly repeated DNA sequence family. *Mol. Biol. Evol.* 11 (3), 384–392. doi: 10.1093/oxfordjournals.molbev.a040114

Zouré, H. G., Norma, M., Tekle, A. H., Amazigo, U. V., Diggle, P. G., Giorgi, E., et al. (2014). The geographic distribution of onchocerciasis in the 20 participating countries of the African Programme for Onchocerciasis Control: (2) pre-control endemicity levels and estimated number infected. *Parasit. Vectors* 7, 326. doi: 10.1186/1756-3305-7-326

Zuchi, A., Prust, L. T., Rocha, A., Araujo, J., da Silva, P. S., Fiorillo, K., et al. (2017). Screening and evaluation of lymphatic filariasis in immigrants from endemic countries residing in a focus where it is considered eliminated in the Southern Region of Brazil: a risk of reemergence? *Acta Trop.* 176, 192–196. doi: 10.1016/j.actatropica.2017.08.010

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read
for greatest visibility
and readership

FAST PUBLICATION
Around 90 days
from submission
to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative,
and constructive
peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers
acknowledged by name
on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: info@frontiersin.org | +41 21 510 17 00

REPRODUCIBILITY OF RESEARCH
Support open data
and methods to enhance
research reproducibility

DIGITAL PUBLISHING
Articles designed
for optimal readership
across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics
track visibility across
digital media

EXTENSIVE PROMOTION
Marketing
and promotion
of impactful research

LOOP RESEARCH NETWORK
Our network
increases your
article's readership