Genetic architecture of transmission stage production and virulence in schistosome parasites

Authors: Winka LE CLEC'H*, Frédéric D. CHEVALIER, Marina McDEW-WHITE, Vinay MENON, Grace-Ann ARYA, , Timothy J.C. ANDERSON

Affiliation: Texas Biomedical Research Institute, P.O. Box 760549, 78245 San Antonio, Texas, USA.

Email addresses:

Winka LE CLEC'H: winkal@txbiomed.org – ORCID: 0000-0002-1111-2492

Frédéric D. CHEVALIER: fcheval@txbiomed.org – ORCID : 0000-0003-2611-8106

Marina McDEW-WHITE: mwhite@txbiomed.org – ORCID: 0000-0001-7821-260X

Vinay MENON: menon34@purdue.edu

Grace-Ann ARYA: apanga@txbiomed.org

Timothy J.C. ANDERSON: tanderso@txbiomed.org – ORCID: 0000-0002-0191-0204

*Corresponding author

ABSTRACT:

Both theory and experimental data from pathogens suggest that the production of transmission stages should be strongly associated with virulence, but the genetic bases of parasite transmission/virulence traits are poorly understood. The blood fluke *Schistosoma mansoni* shows extensive variation in numbers of cercariae larvae shed and in their virulence to infected snail hosts, consistent with expected trade-offs between parasite transmission and virulence. We crossed schistosomes from two populations that differ 8-fold in cercarial shedding and in their virulence to *Biomphalaria glabrata* snail hosts, and determined four-week cercarial shedding profiles in F0, F1 and 376 F2 progeny from two independent crosses in inbred snails. Sequencing and linkage analysis revealed that cercarial production is polygenic and controlled by five QTLs. These QTLs act additively, explaining 28.56% of the phenotypic variation. These results demonstrate that the genetic architecture of key traits relevant to schistosome ecology can be dissected using classical linkage mapping approaches.

KEY WORDS: *Schistosoma* parasite, *Biomphalaria* snails, life history traits, virulence, transmission, genetic crosses, quantitative trait loci (QTL).

INTRODUCTION:

Parasites face a central problem: how to maximize transmission to the next host. This has driven the evolution of a wide variety of lifecycle features to facilitate parasite transmission (Criscione et al., 2020; Vasudevan et al., 2015). However, perhaps the most common transmission strategy is to produce vast numbers of infective stages (Loker and Hofkin, 2015, p. 20). This brute force approach exploits host resources for parasite growth and reproduction, while infective stages may also cause damage as they exit the host to reach the environment (Jensen et al., 2006, p. 200; Le Clec'h et al., 2019). Hence, there is a strong expectation that transmission stage production will result in collateral damage to the host, but also that high levels of host mortality will constrain evolution of very high levels of parasite virulence. There is a large body of theoretical work examining this relationship between transmission stage production and virulence evolution (Alizon et al., 2009; Anderson R.M., 1982; Frank, 1996) and empirical studies provide compelling, but not universal support, for this model (Acevedo et al., 2019; Alizon and Michalakis, 2015). However, there is limited understanding of the genes and genetic architecture underlying transmission stage production and virulence on which selection can act.

Schistosome parasites provide a useful system for examining the genetic basis of transmission/virulence related traits. These parasitic organisms are well suited for genetic studies because (i) the complete lifecycle can be maintained in the laboratory using rodent definitive hosts and freshwater snail intermediate hosts, (ii) parasites have separate sexes which simplifies staging of efficient genetic crosses in the laboratory, (iii) thousands of progeny are produced which provides good statistical power, and (iv) experimental work over the last 75 years has revealed heritable genetic variation in multiple biomedically important traits (Anderson et al., 2018) such as drug resistance (Cioli et al., 1992; Greenberg, 2013; Melman et al., 2009; Mwangi et al., 2014; Valentim et al., 2013),

chronobiology (Théron, 2015; Théron and Combes, 1983, 1988), host specificity (Files and Cram, 1949; Kalbe et al., 2004; Mitta et al., 2017; Rollinson et al., 2001; Theron et al., 2014), and virulence (Davies et al., 2001; Gower and Webster, 2004; Webster et al., 2004). Furthermore, the *Schistosoma mansoni* genome is fully sequenced and assembled (Berriman et al., 2009; Protasio et al., 2012), and a growing molecular toolkit including molecular sexing tools (Chevalier et al., 2016; Gasser et al., 1991), RNAi (Krautz-Peterson et al., 2010), transfection (Mann et al., 2014; Rinaldi et al., 2012), CRISPR (Ittiprasert et al., 2019; Sankaranarayanan et al., 2020; You et al., 2021) and a suite of cell biology tools (Collins and Collins, 2017; Wang et al., 2020; Wendt et al., 2020; Wendt and Collins, 2016) improves our ability to link phenotype with genotype. Furthermore, we can also control the genetics of the snail host by generating inbred snail lines. *B. glabrata* is hermaphrodite, so inbred snail lines can be developed by isolating individual snails and serial inbreeding (Bonner et al., 2012).

In addition of being a tractable model for genetic studies, schistosomes are also devastating human parasites. Three major schistosome parasite species infect over 200 million people in 78 countries (Verjee, 2019; World Health Organization, 2019). These parasites have a complex lifecycle, involving a freshwater snail (intermediate host) and a mammal (definitive host). When parasite eggs are expelled with mammal feces or urine in freshwater, miracidia larvae hatch and actively search for its snail vector. Larvae penetrate the snail head-foot, differentiate into sporocysts that then asexually proliferate to generate daughter sporocysts. These actively consume snail tissue (hepatopancreas and ovotestis), castrating the host snail, and parasite sporocysts can comprise up to 60% of the tissue within infected snails (Le Clec'h et al., 2019). The daughter sporocysts release cercariae, the mammal-infective larval stage of the parasite. Hundreds to thousands of these motile cercariae exit through the snail body wall and are released into freshwater. Exit through the body wall results in leakage of hemolymph

and damage to the snail. This quantitative transmission and virulence related trait – numbers of cercarial produced from the intermediate snail host – is the focus of this paper.

There is strong evidence from laboratory selection experiments that this cercarial shedding is heritable: in just three generations parasites selected for low or high numbers of cercariae showed rapid divergence in this phenotype (Davies et al., 2001; Gower and Webster, 2004). We have described distinctive life-history strategies in laboratory schistosome populations. In two parasite populations originating from Brazil we observed: i) a “boom-bust” strategy characterized by high transmission measured in term of cercarial production (the number of larvae released from the intermediate snail host) and high virulence to the snail intermediate host resulting in short duration infections, compared with ii) a “slow and steady” strategy characterized by low transmission (few cercariae released), resulting in low virulence to the snail host and a long duration of infection (Le Clec’h et al., 2019). The high shedding genotype develops larger sporocyst stages than the low shedding genotype, although this is insufficient to explain the difference in cercarial shedding from these two parasites. These patterns are observed when using the same inbred snail line, ruling out an impact of host genetics. Our central goal was to decipher the genetic architecture of a transmission/virulence-related trait in a biomedically important helminth parasite, to determine whether cercarial shedding is under monogenic or polygenic control, and ultimately to understand the cellular pathways on which selection acts to modulate this trait.

Linkage mapping has previously been used to examine a monogenic drug resistance trait (oxamniquine resistance) in *S. mansoni* (Chevalier et al., 2014; Valentim et al., 2013). Here we extend this approach to examine a life-history trait: we conducted reciprocal genetic crosses between *S. mansoni* individuals from two laboratory parasite populations originating from Brazil: the SmBRE (Low

Shedder - L) population and the SmLE (High Shedder - H) population (Figure 1). Using classical linkage mapping, we showed that cercariae production in *S. mansoni* parasite is polygenic, with additive variation at 5 different QTLs explaining 28.56% of the variation in cercarial production.

Figure 1: Outline of the genetic cross experimental design. We performed two independent reciprocal genetic crosses (i.e crosses A and B, to account for potential sex specific effect(s)) between single genotypes of SmLE-H and SmBRE-L *Schistosoma mansoni* parasite populations. These two populations' exhibit striking differences in term of transmission stage production (i.e. number of cercariae produced). For each parasite generation (i.e F0 parental populations, F1s and F2s progeny), we exposed individual *Biomphalaria glabrata* Bg26 inbred snail to single miracidium from either the SmLE-H or SmBRE-L populations for F0 ($n=192$ /population), or the F1 progeny ($n=288$ /cross), or the F2 progeny ($n=1000$ /cross). For each infected snail and each generation of *S. mansoni* parasites, we measured transmission stage production during 4 weeks of the patent period (week 4 to 7 post-infection). We also evaluated the virulence of these generations of crossed parasites by measuring the daily snail survival during the patent period. After 7.5 weeks post-infection, surviving infected snails were bled and we measured the total laccase-like activity as well as the hemoglobin rate in the collected hemolymph samples.

RESULTS:

Variation in transmission stage production in S. mansoni crosses

We conducted two three-generation (F0 to F2) genetic crosses between high (H) shedding parasite genotypes SmLE-H and low (L) shedding genotypes SmBRE-L. We paired a male SmBRE x a female SmLE (cross A) and a female SmBRE x a male SmLE (cross B) to allow examination of sex specific inheritance. The two *S. mansoni* populations, SmLE-H and SmBRE-L, both originated from Brazil, and show striking difference in cercarial production (Le Clec'h et al., 2019). We used the same inbred *B. glabrata* population (Bg26) in these experiments to minimize the impact of the host genetics, because cercarial shedding can be influenced by snail genotype (Jones-Nelson et al., 2011). Furthermore, we infected snails with single parasite miracidia, so quantitative measure of cercarial production by infected snails can be related to a single parasite genotype.

Using weekly measures of cercarial shedding over 4 weeks, we observed that SmLE (H) population shed 8-fold more cercariae than SmBRE (L) (mean ($\pm$ se) cercariae per shedding: SmBRE (L): 284 ± 19 vs SmLE (H) 2352 ± 113 ; Wilcoxon test, $p < 2.2 \times 10^{-16}$, Figure 2A-B F0; Le Clec'h et al., 2019). However, the infectivity of these two parasite population in Bg26 snails was identical (i.e. 26% of the snails exposed with one miracidia from either SmLE-H or SmBRE-L were infected).

We measured the cercarial production of the two *S. mansoni* F1 populations (i.e. F1A for cross A and F1B for cross B). For both crosses, F1 parasite populations shed intermediate numbers of cercariae compared to F0 (i.e. H and L parental populations) (Kruskal-Wallis test followed by pairwise Wilcoxon Mann-Whitney post-hoc test, $p < 2.2 \times 10^{-16}$) and showed limited variation in shedding number (Figure 2A). Schistosome gender impacted the number of cercariae produced by the F1A parasite population with male genotypes producing significantly less cercariae than female ones (one-

Figure 2: Transmission stage production of two *Schistosoma mansoni* parental populations (i.e. SmLE-H and SmBRE-L) and their progeny (i.e. F1s and F2s). (A) Difference in the average of cercariae produced by SmLE-H and SmBRE-L *S. mansoni* populations during 4 weeks of the patent period (week 4 to 7 post infection). SmLE-H population is shedding more cercariae than SmBRE-L population (Le Clec'h et al., 2019). For both crosses A and B, F1s populations exhibit intermediate phenotype in terms of cercarial production, compared to F0, while the average production of cercariae by F2s encompass both parental and F1s distributions. Parasite populations or generations not connected by the same letter are significantly different (post-hoc test). (B) Difference in the number of cercariae produced by *S. mansoni* parental populations (i.e. SmLE-H and SmBRE-L) and progeny (i.e. F1 and F2) measured by week (week 4 to 7 post infection). (C) Distribution of the cercarial production (mean $\pm$ SE) over the 4 weeks of the patent period for the parents (F0) of the cross A, the F1 parents (F1A) and all the 204 F2A progeny by rank order. (D) Similar distribution plot of the cercarial production for the parents (F0), the F1 parents (F1B) and all the 204 F2B progeny for the cross B by rank order. For both crosses (C-D), L and H F0 parents exhibited striking differences in cercarial production (cross A: Welsh t-test, $p = 0.031$; cross B: Welsh t-test, $p = 0.013$) while F1 showed an intermediate phenotype and F2 encompassed both parental distribution with a gradation from L to H phenotype. NS: No significant difference in cercarial production between the two considered groups, * $p < 0.05$, ** $p \leq 0.02$, *** $p \leq 0.002$.

way ANOVA, $p = 0.0038$, Supplementary figure 1A). We did not see any effect of parasite gender in the F1B progeny (Supplementary figure 1B). Moreover, F1 parasites from cross A produced significantly more cercariae than those from cross B (Wilcoxon test, $p = 0.0048$, Figure 2A-B).

The F2 parasite progeny from both crosses showed extensive variation in numbers of cercariae shed that encompassed the range seen in the two parental distributions (Kruskal-Wallis test followed by pairwise Wilcoxon Mann-Whitney post-hoc test between F0 populations (i.e. SmLE-H and SmBRE-L) and F2s (i.e. F2A and F2B), $p < 2.2 \times 10^{-16}$, Figure 2A F2). In F2s from both crosses, we observed parasites that produced very few cercariae, like the L parent, some that produced cercariae at an intermediate level, like F1s, and some that produced large numbers of cercariae like the H parent (Figure 2C and 2D). F2 progeny from cross A shed more larvae than cross B (Wilcoxon test, $p = 0.0089$, Figure 2A-B F2). The F2 progeny have an average cercarial shedding profile across weeks that mimic the L parental profile with a peak in cercarial production arising on the second shedding. However, the average intensity of cercarial production is closer to the H parent (Figure 2A-B). We observed a significant impact of schistosome gender in cross B (Kruskal-Wallis test, $p=0.041$, Supplementary figure 1C) but not in cross A (Kruskal-Wallis test, $p=0.719$, Supplementary figure 1D) for the F2s with male genotypes producing significantly less cercariae than female ones.

Linkage analysis of transmission stage production in S. mansoni parasites

We sequenced the whole genome (363 Mb) from parents and the exome (15 Mb) of F1s and F2s progeny from each cross (Figure 1). We found 10,543 (cross A) and 8,779 (cross B) SNPs fixed for alternative alleles (i.e. fully informative markers) in the exome. A classical linkage analysis using as phenotype the average quantities of cercariae per snail (i.e. schistosome genotype) revealed three

major quantitative trait loci (QTL) involved in transmission stage production: on chr. 1 (LOD= 5.63), chr. 3 (LOD= 8.16) and chr. 5 (LOD= 6.37) (Figure 3A). This finding is consistent with the minimum number of loci calculated using the Castle-Wright Estimator (Castle, 1921; Cockerham, 1986; Hedrick, 1983; Lande, 1981; Wright and Morton, 1968), which estimates the minimum number of loci determining trait variation from patterns of segregating phenotypic variation. This estimated a minimum of 3.03 loci involved for cross A and 1.85 for cross B. Gender taken as a covariate did not impact the result of the linkage analysis.

We used a two-dimensional genome scan to search for additional contributing QTLs and to identify interactions among loci. We found two minor QTLs on chromosome 2 (LOD= 3.87) and 4 (LOD= 3.32) which contributed significantly to explain the phenotype variance. The two-QTL scan demonstrated that all five QTLs act additively: we found no evidence for epistasis between the loci (Table 1). Together, these 5 QTLs explained 28.56% of the variation in cercarial production, with a full LOD score of 29.79 (Table 2). The three major QTLs explained most of the phenotype: chr. 1 = 8.10% ($p = 5.49 \times 10^{-10}$), chr. 3 = 6.10% ($p = 8.49 \times 10^{-8}$), chr. 5 = 5.57% ($p = 3.31 \times 10^{-7}$). The two minor QTLs

Table 1: Summary table of the two-dimensional, two-QTL genome scan identifying QTL interactions (additive or epistatic effects) between the QTL involved in cercarial production in *Schistosoma mansoni* parasite. LOD values are statistically significant when they are above the 5% threshold LOD for epistatic (interaction) and/or additive effect. We have demonstrated here additive interaction of various loci involved in transmission stage production but no epistatic interaction.

Pairs of loci	Position locus 1	Position locus 2	LOD interaction (epistatis)	LOD additive
5% threshold	-	-	5.68	5.84
Chr1:Chr2	43.84	45.94	0.611	10.48
Chr1:Chr3	43.84	5.44	0.186	12.40
Chr1:Chr4	43.84	28.95	1.020	9.05
Chr1:Chr5	43.84	21.56	1.367	11.88
Chr2:Chr4	45.94	28.90	0.738	7.48
Chr2:Chr5	45.94	21.56	1.244	9.60
Chr3:Chr5	5.44	21.56	2.587	10.96
Chr4:Chr5	28.90	21.56	0.622	9.20

explained less phenotypic variation: chr. 2 = 3.80% ($p = 3.38 \times 10^{-5}$) and chr. 4 = 3.52% ($p = 7.06 \times 10^{-5}$) but these contributions are highly significant.

Figure 3. Linkage analysis of the transmission stage production in *S. mansoni* parasites and QTLs interactions. (A) Linkage analysis between the average cercarial production phenotype and F2 *S. mansoni* genotypes (combined crosses A and B) demonstrated that transmission stage production is a polygenic trait controlled by three major QTLs with statistically significant LOD: on chr. 1 (LOD= 5.63), on chr. 2 (LOD= 8.16) and on chr. 5 (LOD= 6.37). The blue dotted line represents the 1,000 permutation threshold. (B) Genetic architecture of cercarial production is influenced by the shedding week (first shedding week corresponds to week 4 post-infection), demonstrating a sequential pattern of QTL emergence involved in transmission stage production. On shedding week 1, none of the QTLs identified pass the permutation threshold, while on shedding week 2, cercarial production is linked to QTL on chr. 5. On shedding week 3, the main QTL linked to cercarial production is on chr. 3. Finally, on shedding week 4, the QTL on chr.1 predominates, along with QTLs on chr. 2 and chr. 3.

Table 2: QTLs interactions modeling (additive model) and estimation of the proportion of the cercarial production variance (%var) explained by each of the loci involved in the phenotype. All together, the 5 QTLs explained 28.56% of variation in the transmission stage production in *S. mansoni* parasites.

Loci	LOD	%var	P value
Full additive model (all 5 QTLs)	29.8	28.56	0
Chr1	9.52	8.10	5.49x10⁻¹⁰
Chr2	4.6	3.80	3.38x10⁻⁵
Chr3	7.26	6.10	8.49x10⁻⁸
Chr4	4.26	3.52	7.06x10⁻⁵
Chr5	6.66	5.57	3.31x10⁻⁷

Genetic control of shedding varies across time

When we conducted the linkage analysis using the cercarial quantities produced each week rather than the average across all four weeks (i.e. from week 4 to week 7 post-exposure, see Figure 1), we observed a sequential pattern of QTL emergence (Figure 3B, Supplementary figure 2). On shedding week 1, none of the identified QTLs passed the permutation threshold (Figure 3B, Supplementary figure 2A), while on shedding week 2, the QTL on chr. 5 predominated and a QTL on chr. 3 had arisen (Figure 3B, Supplementary figure 2B). On shedding week 3, the main QTL was on chr. 3 (Figure 3B, Supplementary figure 2C). Finally, on shedding week 4, the QTL on chr. 1 predominated, along with QTLs on chr. 2 and 3 (Figure 3B, Supplementary figure 2D). Hence, the parasite genes that determine cercarial production vary across time during the snail infection.

Allelic inheritance and impact on cercarial production in *S. mansoni* parasites

For each QTL, we determined the interactions between alleles from L and H shedding parents by analyzing the average production of cercariae per genotype (i.e. LL: homozygous for the low shedding allele, LH: heterozygous and HH: homozygous for the high shedding allele; Figure 4). Our analysis

Figure 4. Inheritance of transmission stage production alleles. Impact of the parasite genotype on the average number of cercariae produced for each major and minor QTL linked to transmission stage production. On chr. 1, cercarial production is not significantly different when parasites are homozygous for the “low shedding” allele (i.e. LL) or heterozygous (i.e. LH) but only when parasites are homozygous for the “high shedding” allele (i.e. HH). High shedding allele (H) is recessive and low shedding allele (L) is dominant. On chr. 2, cercarial production is significantly different for all the three parasite genotypes encountered (i.e. LL, LH and HH). High and Low shedding alleles act co-dominantly. On chr. 3, we observed the same results as exhibited for chr. 2. For this locus, High and Low shedding alleles also act co-dominantly. For the locus located on chr. 4, we encounter the same scenario as for the locus on chr. 1: the cercarial production is not significantly different for parasites exhibited LL or LH genotypes but for the HH genotypes. H allele is recessive and L dominant. Finally, on chr. 5, we observed similar results as already shown for chr. 2 and 3, where the cercarial production is significantly different whatever the parasite genotype. L and H alleles are co-dominant for this locus. Parasite genotypes (i.e. LL, LH, HH) not connected by the same letter for a given locus are significantly different (post-hoc test).

showed codominance for the chr. 2, 3 and 5 QTLs: F2 progeny with LH genotype produced more cercariae than LL genotype but less HH genotype (Figure 4). However, for the chr. 1 and 4 QTLs showed patterns consistent with recessive inheritance. At these loci, F2 progeny with HH genotype produced more cercariae (pairwise comparison using Wilcoxon rank sum test; chr. 1: HH vs. LH: $p = 4.2 \times 10^{-5}$ and

HH vs. LL: $p = 8.2 \times 10^{-6}$; chr. 4: HH vs. LH, $p = 0.0109$ and HH vs. LL: $p = 0.0034$) than the other genotypes (i.e. LL and LH; pairwise comparison using Wilcoxon rank sum test; chr. 1: LL vs. LH, $p = 0.22$; chr. 4: LL vs. LH, $p = 0.56$; Figure 4).

Candidate genes controlling transmission stage production in S. mansoni parasite

We listed the genes located within the 1.8 LOD-support interval of each QTL and expressed either in daughter sporocysts or in cercariae. We identified 314 genes on chr. 1, 87 on chr. 2, 31 on chr. 3, 76 on chr. 4 and 86 on chr. 5 (Table 3 and Supplementary Table 1). We then prioritized genes using an objective index ("genoscore", see Material & Methods section), that accounts for the genotypes of the parents (fix alternative alleles), the presence of non-synonymous mutations and their impact on the protein predicted by snpEff (see Material & Methods section), the CDS length and the distance from the QTL peak. Using this objective approach, we identified the three most likely candidate genes within each QTL that may be involved in transmission stage production (Table 4).

The genes prioritization on each of the three major QTLs revealed three compelling candidates: (i) peptidylprolyl isomerase (chr. 1) which regulates many biological processes, including intracellular signaling, transcription, inflammation, immunomodulation and apoptosis (Ramachandran et al., 2014;

Table 3: Summary table of the total number of genes located within the 1.8 LOD-support interval of each QTL peak, and expressed into the different parasite stages: daughter sporocysts only, cercariae only, both stages or not expressed at all.

Loci	Total number of genes	Genes expressed in daughter sporocysts only	Genes expressed in cercariae only	Genes expressed in both stages	Genes not expressed
Chr1	329	297	301	284	15
Chr2	94	87	83	78	2
Chr3	32	28	28	25	1
Chr4	79	70	72	66	3
Chr5	94	78	82	74	8

Wei et al., 2013), (ii) G-protein coupled receptor kinase 2 (chr. 3) which is involved in cell migration (Randazzo et al., 2000) and (iii) a protein with unknown function (chr. 5) which suggests schistosome specific factor yet to be characterized. On chr. 5, another candidate, the Calmodulin IQ domain protein, is of interest: this protein is a calcium sensor and can stimulate changes in the actin cytoskeleton mediated by proteins such as myosin (Houdusse and Cohen, 1996).

In the two minor QTLs, the top candidate genes encode a nucleoporin (chr. 2) and a protein of unknown function (chr. 4). The second best candidate gene on chr. 4 is also of interest: it encodes a LYAR, a cell growth regulating nucleolar protein.

Linkage analysis of snail physiological trait associated with transmission stage production

Laccase-like activity and hemoglobin rate in the *Biomphalaria* snail hemolymph were correlated (Supplementary figure 3C and 3F). These metrics provide proxies to evaluate snail health and the impact of schistosome infection (Le Clec'h et al., 2016; Le Clec'h et al., 2019). Both parameters were negatively correlated with F2 cercarial production (laccase-like activity: Pearson's correlation test, $p = 4.31 \times 10^{-15}$, correlation coefficient = -0.56; Hemoglobin rate: Pearson's correlation test, $p = 8.014 \times 10^{-12}$, correlation coefficient = -0.50; Supplementary figure 3D and 3E). There was no correlation with the F1 cercariae production (laccase-like activity: Pearson's correlation test, $p = 0.426$, correlation coefficient = -0.09; Hemoglobin rate: Pearson's correlation test, $p = 0.661$, correlation coefficient = -0.05; Supplementary figure 3A and 3B). This was expected because F1s were heterozygous for markers from each parent at all five QTLs and show limited variation in cercarial production.

227 **Table 4: Summary table of the 3 best candidate genes for each locus involved in cercarial production in the parasite *S. mansoni* based on the**
228 **computed genoscore CDS (see Materials & Methods) and their relative positions to the maximum peak. NA: gene not captured by the *S. mansoni* baits**
229 **(exome capture). In bold: selected best candidate genes based on their annotations (GFF or HHsearch).**

Chromosome	Gene ID	GFF annotation	HHsearch annotation	Max. LOD	Genoscore
Chr1	Smp_152850	Tetratricopeptide-like helical domain superfamily	Putative 70 kda peptidylprolyl isomerase PFL2275c	5.63	1.646
	Smp_057230	Aminoacyl-tRNA synthetase, class II	Seryl-tRNA synthetase (SerRS)	NA	1.286
	Smp_246820	<i>No annotation available</i>	hypothetical protein TTHA1254	4.05	1.114
Chr2	Smp_347740	Coiled-coil domain-containing protein 39	Nucleoporin_FG2	3.86	7.657
	Smp_328700	<i>No annotation available</i>	Engineered transmembrane domain variant	NA	2.153
	Smp_170430	FAD dependent oxidoreductase	D-aminoacid oxidase, N-terminal domain	2.81	1.491
Chr3	Smp_343200	Pleckstrin homology domain	G-protein coupled receptor kinase 2 (beta-adrenergic receptor kinase 1)	8.16	1.653
	Smp_190770	Lanthionine synthetase C-like	Nisin biosynthesis protein NisC	5.41	1.426
	Smp_138060	<i>No annotation available</i>	Cationic trypsin (e.c.3.4.21.4)	7.66	1.227
Chr4	Smp_313330	<i>No annotation available</i>	<i>No annotation available</i>	3.35	1.732
	Smp_150840	Zinc finger, C2H2, LYAR-type	60S ribosomal protein L8, 60S	2.47	1.502
	Smp_334250	Epithelial sodium channel	Acid-sensing ion channel 1	3.42	1.257
Chr5	Smp_330680	<i>No annotation available</i>	<i>No annotation available</i>	NA	6.997
	Smp_330700	<i>No annotation available</i>	<i>No annotation available</i>	NA	3.784
	Smp_330760	IQ motif, EF-hand binding site	Calmodulin, IQ domain-containing protein G	4.50	3.631

232 We performed a linkage analysis to investigate parasite genes linked to laccase-like activity and
233 the hemoglobin rate in the snail host hemolymph. We did not find significant QTLs across *S. mansoni*
234 genome linked to these snail hemolymph phenotypes.

235

DISCUSSION:

How many loci are involved in transmission/virulence?

We determined the genetic architecture of transmission stage production trait in schistosome parasites using classical linkage mapping. We showed that transmission/virulence traits in schistosome are polygenic, involving three major and 2 more minor loci. These results are consistent with the idea that relatively small numbers of genes are involved in strongly selected traits such as virulence/transmission. The obvious caveat here is that loci of small effect may be difficult to identify even in well powered linkage analyses.

Three other studies have used linkage mapping to dissect the genetic basis of parasite transmission/virulence trait (Table 5). In the protozoan parasite *Toxoplasma gondii*, four sets of genetic crosses were used to identify the genetic basis of acute virulence between different clonal lineages from North America (Behnke et al., 2014). These crosses identified several QTLs linked to acute virulence and resulted in the identification of a set of polymorphic genes in the GTPase serine/threonine kinase family in secretory organelles called rhoptries (Behnke et al., 2016). In another protozoan model, the rodent malaria parasite (*Plasmodium yoelii yoelii*), Pattaradilokrat et al. (2009) and Otsuki et al. (2009) showed that just one locus on chr. 13 was linked to parasite growth rate and host virulence. Plant parasite nematode system provides a directly comparable macroparasite system in which virulence has been examined using linkage mapping. In the root-knot nematode *Meloidogyne hapla*, crosses between parasites that differ in ability to produce galls on the roots of the ornamental nightshade *Solanum bulbocastanum* identified three QTLs linked to nematode transmission success (Thomas and Williamson, 2013).

257 In all these three cases, relatively few loci control transmission/virulence traits, as we see in
 258 schistosome parasites (Table 5). This question of “few loci or genes” versus “many loci or genes”
 259 controlling key life-history traits is fundamental to our understanding of adaptation in nature. As

Table 5: Do transmission/virulence traits tend to be oligogenic (i.e. controlled by few genes)? Summary table of the number of loci and genes involved in transmission/virulence or growth life-history traits for various eukaryotic parasite systems (reviewed from the literature).

Parasite system	Phenotypes	Number of loci involved (Chrom. #)	Number of gene(s) involved (Gene name)	Percentage of phenotypic variations explained by the loci	References
<i>Toxoplasma gondii</i>	Acute virulence	1 (XII)	1 (ROP5)	-	Behnke et al., 2011
	Acute virulence	3 (VIIa/VIIb/XII)	3 (ROP18/ROP16/ROP5)	-	Reese et al., 2011; Saeij et al., 2006
	Acute virulence	2 (VIIa/Ia)	1 (ROP18)	-	Fentress et al., 2010; Taylor et al., 2006
	Growth	4 (VIIa/XI/XII/Ia)	-	-	Yamamoto et al., 2011
	Migration	1 (VIIa)	-	-	Taylor et al., 2006
	Acute virulence	1 (XII)	1 (ROP5)	-	Behnke et al., 2015
<i>Plasmodium yoelii yoelii</i> (rodent malaria parasite)	Growth rate	1 (XIII)	1 (<i>pyebl</i>)	-	Otsuki et al., 2009; Pattaradilokrat et al., 2009
<i>Meloidogyne hapla</i> (plant pathogen nematode)	Egg mass number	2 (VII/XIII)	-	-	Thomas and Williamson, 2013
	Total eggs among F2 lines	2 (IV/VII)	-	-	Thomas and Williamson, 2013
<i>Schistosoma mansoni</i>	Transmission stage production/Virulence	5 (3 major: I/III/V and 2 minor: II/IV)	5 potential best candidate genes (peptidylprolyl isomerase/ G-protein coupled receptor kinase / Calmodulin and Nucleoporin_FG2/ LYAR cell growth regulating nucleolar protein)	28.56%	<i>Present study</i>

demonstrated by an increasing number of studies, few large-effect loci (and subsequently genes) have been shown to drive rapid adaptation (reviewed in Messer et al., 2016).

How do parasite genes impact numbers of cercariae shed?

Genes may impact numbers of cercariae shed from infected snails by (i) influencing growth and differentiation of sporocysts, (ii) changing the balance of investment between cercarial production and daughter sporocyst production, (iii) influencing the ability of cercariae to escape from the snail host (Figure 5). Two lines of evidence strongly suggest that the genes determining cercarial shedding may act on at least different processes.

First, we previously compared sporocyst size and growth kinetics in the parental populations using qPCR. SmLE-H sporocysts comprise on average 47% of cells within infected snails, while SmBRE-L sporocysts comprise 25% of cells within infected snails (Le Clec'h et al., 2019). However, this is not sufficient to fully explain the difference in cercariae produced by these two populations, because the SmBRE-L infected snails shed significantly fewer cercariae than predicted from qPCR measures of sporocysts cells in infected snails. These results are consistent with independent action of genes to determine sporocyst size and cercarial production.

Second, when we dissected the genetic architecture of cercarial production over the 4-week patent period of the infection, we observed sequential emergence of QTLs. This is also consistent with the idea that genes influencing cercarial shedding may act on different processes to determine this cercarial production. This sequential emergence of loci could also explain the evolution of sporocysts kinetics trajectories observed in parental population (Le Clec'h et al., 2019). Moreover, we highlighted that the high shedding alleles acted co-dominantly with the low shedding alleles at QTLs on chr. 2, chr. 3 and chr. 5 but were recessive at QTLs on chr. 1 and chr. 4. These differences in allelic inheritance

could also support the existence of different mechanisms involved in the modulation of transmission stage production and virulence: co-dominant loci would involve alleles with dose effect while recessive loci would involve loss-of-function alleles. In the next section we evaluate potential candidate genes and speculate at which stage of the parasite these may play a role.

Candidate genes underlying transmission/virulence traits

The first QTL linked to cercarial production appears on the second shedding week and is located on chr. 5. The three leading candidate genes under this QTL include two genes encoding proteins of unknown function and a gene encoding a calmodulin (CAM) IQ domain protein. The proteins of unknown function are of high interest as they could reveal new mechanisms involved in transmission/virulence and will deserve further attention. CAM IQ domain proteins belong to calcium sensors and can stimulate changes in the actin cytoskeleton mediated by proteins such as myosin (Houdusse and Cohen, 1996). In schistosome, the calcium sensor (CAM IQ domain protein) could modulate the differentiation of daughter sporocysts cells into cercariae and induce the cercarial shedding early in the patent period. Calcium binding proteins are known to play an important role in amoebic pathogenesis and are essential for the *Entamoeba histolytica* parasite growth (Sahoo et al., 2004), while in *Toxoplasma gondii* parasite, CAM-like proteins are essential and contribute to regulate parasite motility and host cell invasion (Long et al., 2017).

During the third shedding week, the major QTL is on chr. 3, where the leading candidate gene is a G-protein coupled receptor kinase 2. In *Entamoeba histolytica* G-protein are involved in pathogenesis-related cellular processes, such as migration, invasion, phagocytosis and evasion of the

host immune response by surface receptor capping (Bosch and Siderovski, 2013). In *S. mansoni*, this G-protein coupled receptor kinase 2 may stimulate sporocyst expansion or renewal into snail host tissues (i.e. new generation of daughter sporocysts) while escaping host immune defenses, to maintain the parasitic infection and eventually, cercarial production.

During the fourth shedding week, QTLs on chr. 1, chr. 2 and chr. 3 are involved in transmission stage production. The leading candidate gene on chr. 1 is a peptidylprolyl isomerase (PPIase). Most PPIases characterized in parasites belong to the cyclophilin family (Ünal and Steinert, 2014) and are strong immunomodulatory proteins (Wang and Heitman, 2005). In the apicomplexan *Theileria*, a secreted prolyl isomerase modulates host leukocyte transformation (Marsolier et al., 2015); similar immunomodulatory peptidylprolyl isomerases have been demonstrated in two other apicomplexan parasites, *Toxoplasma gondii* and *Neospora caninum* (Tuo et al., 2005). PPIases have been shown to be primary actors of host-parasite interactions and are certainly essential for the development and differentiation of parasitic protozoans (e.g. *Leishmania* parasites), which show a high degree of plasticity in their cellular organization and metabolic status during their infection cycles (Yau et al., 2010). In our schistosome model, PPIase may modulate snail host immune response to infection, while maintaining proliferation and viability of sporocysts cells. The best candidate gene under the chr. 2 QTL is a nucleoporin, protein that is essential to microtubule organization and dynamics during mitosis. In the parasite *Plasmodium falciparum*, nucleoporins are essential for parasite proliferation in human erythrocytes (Dahan-Pasternak et al., 2013). Nucleoporins are also essential in the transport of mRNA from the nucleus to the cytoplasm after transcription and are involved in cell migration (Chatel and Fahrenkrog, 2012). In schistosome, these proteins might trigger the sporocysts clonal proliferation and cercariae production.

325 The QTL on chr. 4, revealed by the two-dimensional genome scan, did not pass the permutation
326 threshold at any of the 4-week patent period of the infection. Therefore, this QTL could have a minor
327 but continuous effect throughout the patent period. Interestingly, the leading candidate genes under

Figure 5. Potential mechanisms involved in production of schistosome transmission stage. Our genetic analysis of transmission stage production and virulence in *S. mansoni* parasites revealed that this complex trait is controlled by 5 loci with potentially different and complementary mechanisms involved to induce and maintain cercarial production over time. This could involve (i) first the differentiation of sporocysts cells into cercariae through the Calmodulin calcium sensor (chr. 5) (ii) then the maintenance of the clonal reproduction of sporocysts and the continuous cercarial production through the G protein (chr. 3) and the Nucleoporin (chr. 2), and finally (iii) the modulation of the snail immune response to protect parasitic cells through the PPlase (chr. 1) and the G protein (chr. 3). The LYAR cell growth protein (chr. 4) could also be involved all along the patent period to regulate the growth and multiplication of sporocysts cells and the development of cercariae. The proteins of unknown function are also of high interest: they could reveal new mechanisms involved in transmission/virulence and will deserve further attention.

328 this QTL encode protein of unknown function followed by a LYAR cell growth regulating nucleolar
329 protein. Proteins of unknown function are again of high interest and will deserve further attention. In
330 addition, the LYAR protein, required for cell proliferation (Hui Li et al., 2009) and highly present in early

embryos of mammals (Su et al., 1993), could also be involved all along the patent period to regulate the growth and multiplication of schistosome sporocysts cells and the development of cercariae.

Prospects for functional analysis

Figure 5 summarizes potential involvement of candidate genes in cercarial shedding in *S. mansoni*. Functional validation is required to investigate involvement of these candidate genes. While RNAi methods are available for gene knockdown in adult worms (Krautz-Peterson et al., 2010), delivery of RNA to the sporocysts within the snail is challenging. We have tried several approaches to induce RNAi in daughter sporocysts including injection of *ds* or *si*RNA into the snail hemolymph/tissues, but have so far not been able to knockdown genes in this stage. CRISPR provide a possible alternative approach but is still at an early stage of development for *S. mansoni* (Ittiprasert et al., 2019; Sankaranarayanan et al., 2020; You et al., 2021). Once effective gene manipulation approaches are available for late sporocysts we will be able to directly examine the involvement of our leading candidate genes.

Stem cells play a central role in schistosome development and reproduction. Wang and colleagues (2018) have demonstrated that primary sporocysts containing different cellular types including stem cells. In our model, we suspect that SmLE-H and SmBRE-L sporocysts may exhibit different cellular trajectories, with differences in development of cell that differentiate to generate cercariae and those that give birth to the next generation of daughter sporocysts (Collins, 3rd et al., 2011). Advances in our understanding of stem-cell differentiation of *S. mansoni* within the molluscan host coupled with single-cell transcriptomics of daughter sporocysts from our two populations provides an alternative approach to investigate the molecular basis of these transmission-related developmental differences at the cellular and molecular levels (Collins, 3rd et al., 2011; Wang et al.,

353 2018, 2013). Functional and cell biology characterization of this key life history trait underlying
354 transmission and virulence will be the focus of future work on this system.

355

MATERIALS & METHODS:

Ethics statement

This study was performed in accordance with the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. The protocol was approved by the Institutional Animal Care and Use Committee of Texas Biomedical Research Institute (permit number: 1419-MA).

Biomphalaria glabrata snails and Schistosoma mansoni parasites

Uninfected inbred albino *Biomphalaria glabrata* snails (line Bg26 derived from 13-16-R1 line; Bonner et al., 2012) were reared in 10-gallon aquaria containing aerated freshwater at 26-28°C on a 12L-12D photocycle and fed *ad libitum* on green leaf lettuce. All snails used in this study had a shell diameter between 8 and 10 mm, as snail size can influence cercarial production (Gérard C., Moné H., 1993; Tavalire et al., 2016). For all the experiments presented in this study, we used inbred snails to minimize the impact of snail host genetic background on the parasite life history traits (Le Clec'h et al., 2019).

The SmLE-H schistosome (*Schistosoma mansoni*) population (high shedder, H) was originally obtained from an infected patient in Belo Horizonte (Minas Gerais, Brazil) in 1965 and has since been maintained in laboratory (Lewis et al., 1986), using *B. glabrata* NMRI and Bg26 population as intermediate host and Syrian golden hamster (*Mesocricetus auratus*) as definitive hosts. The SmBRE schistosome population (low shedder, L) was sampled in 1975 from Recife (East Brazil) (Theron et al., 2014) and has been maintained in the laboratory in its sympatric albino Brazilian snail host (BgBRE) using hamsters or mice as the definitive host.

Genetic crosses between high shedder and low shedder *S. mansoni* parasites

Schistosomes are sex separated parasites that simplifies crosses (Valentim et al., 2013). In order to identify QTL(s) controlling the number of cercariae produced by schistosome parasites, we conducted crosses between parasites from our high shedder (SmLE-H) and low shedder (SmbRE-L) populations (Le Clec'h et al., 2019). We performed two reciprocal genetic crosses (cross A: male SmbRE (L) x female SmLE (H); cross B: female SmbRE (L) x male SmLE (H)) to i) replicate our cross experiment and ii) to test for the potential influence of sex chromosomes on the transmission phenotype (Figure 1).

Design of the crosses is summarized in Figure 1. To obtain the parental (F0) parasite generation we exposed individual snails (192 per parasite populations) to single miracidia in 24 well-plates overnight. Exposed snails were then maintained in trays (48 per tray) for 4 weeks as described in Le Clec'h *et al.* (2019). At four weeks post-exposure, we placed each snail in a well of a 24 well-plate, added in 1 mL freshwater to each well and placed the plate under artificial light for 2 h to induce cercarial shedding. Shedding was conducted late morning to early afternoon every week from week 4 to week 7 post-infection. To track cercarial production of individual snails over the 4 weeks of the patent period, we isolated each infected snail in a uniquely labeled 100 mL glass beaker filled with ~50 mL freshwater at the first shedding, fed them *ad libitum* with fresh lettuce and kept them in the dark during the 4 weeks (except when shedding was induced). We determine parasite gender by PCR (Chevalier et al., 2016) on cercariae collected from week 4, for the two schistosome populations (H and L). One female and one male were randomly selected for producing the next generation. At week 5, cercarial shedding was induced and 2 female golden Syrian hamsters per cross were exposed to 250 cercariae of the female genotype and 250 cercariae of the male genotype. For cross A, the male genotype was SmbRE-L and the female genotype was SmLE-H. For cross B, the male genotype was

SmLE-H and the female genotype was SmBRE-L. After 45 days, we euthanized and perfused the hamsters to recover the F0 adult male and female worms for DNA extraction and sequencing. We also collected the livers containing the eggs.

We applied the same procedures for the F1 generation. We hatched F1 miracidia from eggs recovered from hamster livers following Le Clec'h *et al.* (2019). For each cross, we exposed individual Bg26 snails (288 per cross) to single F1 miracidia. At week 4, we isolated infected snails in individual beakers for tracking individual shedding. We collected cercariae of the first shedding for parasite gender determination. At week 5, we infected 4 female hamsters per cross with 250 cercariae of the females and the male genotypes of the selected F1s. We euthanized and perfused hamsters 45 days post infection to recover F1 adult worms and collected livers containing the F2 eggs.

After hatching F2 miracidia from eggs recovered from hamster livers, we exposed 2,000 individual Bg26 snails (1,000 per cross) to single miracidia. After 4 weeks, we isolated the infected snails in individual beakers as above to track individual shedding. We collected all the cercariae of a snail from the first or the second shedding in individual microtubes for further DNA extractions.

Life history traits measured on S. mansoni parasites and virulence toward its snail hosts

We measured larval output (i.e. cercarial production) of F0, F1 and F2 parasites. We used the same inbred snail population for all infections to minimize the impact of host snail genetic variation (Figure 1). All snails were infected with single miracidia to allow examination of cercarial shedding from single parasite genotypes. We also measured the impact of these parasitic infections on the snail host by quantifying snail survival and physiological responses (i.e. laccase-like activity and hemoglobin rate in

the hemolymph) at end point (Figure 1). The phenotypic datasets are available on Zenodo (DOI: 10.5281/zenodo.4383248).

a. S. mansoni cercarial production over time

We placed infected snails in 1 mL freshwater in a 24 well plate under artificial light (as described above) every week for 4 weeks (week 4 to 7). After replacing the snail back to its beaker, we sampled three 10 (for the high shedder parasites) or 100 μ L (for the low shedder parasites) aliquots of each well, added 20 μ L of 20X normal saline and counted the immobilized cercariae under a microscope. We multiplied the mean of the triplicated measurement by the dilution factor (10) to determine the number of cercariae produced by each infected snail. We assessed cercarial production every week from week 4 to 7 post-exposure in snails infected with F0, F1 and F2 parasites.

b. Snail physiological response to parasitic infection

During week 7, 3 days after the last cercarial shedding, we collected hemolymph from all surviving snails infected with F0 parental populations (i.e. SmLE-H and SmBRE-L), F1s and F2s (Le Clec'h et al., 2016). We measured both laccase-like activity and the hemoglobin (protein carrying oxygen in *B. glabrata* hemolymph) rates in the hemolymph of each infected snail (Le Clec'h et al., 2019).

Estimation of the minimum number of loci influencing transmission stage production

Castle and Wright proposed an estimate n_e of the minimum effective number of genetic factors explaining trait segregation in crosses between two lines based on the phenotypic mean and variance (Castle, 1921; Wright and Morton, 1968).

$$n_e = \frac{(\overline{P_1} - \overline{P_2})^2}{8(Var(F2) - Var(F1))}$$

where P is the phenotypic mean of the parents (i.e. P_1 for SmLE-H and P_2 for SmBRE-L), $Var(F_2)$ is the phenotypic variance of the F2 population progeny and $Var(F_1)$ is the phenotypic variance of the F1 population. We calculated n_e for each cross independently.

Whole genome and exome sequencing of schistosome parasite

We sequenced genomes from F0 parent worms, and exomes from F1 worms used to generate the F2 and from 188 F2s cercariae for both cross A and cross B.

a. gDNA extraction

We extracted gDNA from F0 and F1 worms and from F2 cercariae using the Blood and Tissue kit (Qiagen), following the manufacturer protocol, with minor modifications. We homogenized worms in DNA extraction kit lysis buffer using sterile micro pestles. For F2 cercariae, frozen samples were thawed at 4°C, centrifuged 5 minutes at x300 g to pellet the cercariae. We removed the water supernatant, added lysis buffer to the pellet, and vortexed to homogenize. We incubated at 56°C worm and cercariae samples for 2 and 1 hour, respectively. We recovered gDNA in 200 μ L of elution buffer. We quantified the worm samples gDNA using the Qubit dsDNA HS Assay Kit (Invitrogen) while F2 cercarial DNA was directly whole genome amplified to provide sufficient DNA for exome capture.

b. Whole genome amplification (WGA) of F2 cercarial samples

We performed WGA on each F2 gDNA cercarial sample using the Illustra GenomiPhi V2 DNA Amplification kit (GE Healthcare Life Sciences, USA) according to Le Clec'h *et al.* (2018). We quantified the WGA DNA using the Qubit dsDNA HS Assay Kit (Invitrogen).

c. Whole genome and exome library preparation and sequencing

We prepared whole genome libraries of F0s using the KAPA HyperPlus kit (KAPA Biosystem) according to the manufacturer's protocol. For each F0 library, we sheared 500 ng of gDNA by adaptive focused acoustics (Duty factor: 10%; Peak Incident Power: 175; Cycles per Burst: 200; Duration: 180 seconds) in AFA tubes, using a Covaris S220 instrument with SonoLab software version 7 (Covaris, Inc., USA), to recover fragmented DNA between 150-200 bp. We used 6 PCR cycles for post-ligation library amplification.

We captured F1 and F2 *S. mansoni* exomes using the SureSelect^{XT2} Target Enrichment System (Agilent). The design of the custom baits used to capture the *S. mansoni* exome (SureSelect design ID: S0398493) is described in Chevalier *et al.* (2014) and exome capture methodology follows Le Clec'h *et al.* (2018).

We sequenced the libraries on a HiSeq 2500 sequencer (Illumina) using 100 bp pair-end reads. On each sequencing lane, we either pooled 32 exome capture libraries or 2 whole genome libraries. Raw sequence data are available at the NCBI Sequence Read Archive under accession numbers PRJNA667697 (reviewer link: <https://dataview.ncbi.nlm.nih.gov/object/PRJNA667697?reviewer=bo7f89g1actpms73hb12r2hvn8>).

Bioinformatic analysis

Jupyter notebook and scripts used for processing the sequencing data and mapping QTLs are available on Github (<https://github.com/fdchevalier/Cercarial-production>).

a. Sequence analysis and variant calling

We aligned the sequencing data against the *S. mansoni* reference genome (schistosoma_mansoni.PRJEA36577.WBPS14) using BWA (v0.7.17) (Li and Durbin, 2009) and SAMtools

(v1.10) (Heng Li et al., 2009). We used GATK (v4.1.8) (DePristo et al., 2011; McKenna et al., 2010) to mark PCR duplicates and recalibrate base scores. We used the HaplotypeCaller module of GATK to call variants (SNP/indel) and the GenotypeGVCFs module to perform a joint genotyping on each chromosome or unasssembled scaffolds. We merged VCF files using the MergeVcfs module. We annotated the variants using snpEff (v.4.3.1t) (Cingolani et al., 2012). All these steps were automatized using snakemake (v5.14.0) (Köster and Rahmann, 2018).

b. In silico schistosome sexing

Sexual determination in schistosome relies on Z and W chromosomes: females are ZW while males are ZZ. We can already determine sex molecularly by PCR using sex specific markers (Chevalier et al., 2016). We decided to take advantage of the sequencing data to perform *in silico* sexing by comparing the read depth along the Z chromosome. The Z chromosome carries a Z-linked region (located between 11 and 44 Mb) which never recombines with the W chromosome contrariwise to the rest of the Z chromosome (pseudo-autosomal region). While the Z-linked region in males (ZZ) will display the same read depth as the pseudo-autosomal region, the Z-linked region in females (ZW) will display only half of the read depth of the pseudo-autosomal region (Supplementary Figure 4). We can therefore determine a read depth ratio between the Z-linked and pseudo-autosomal regions: a ratio around 1 will correspond to a male carrying two Z chromosomes while a ratio around 0.5 will correspond to a female carrying only one Z chromosome. We computed the read depth at each base of the chromosome Z using SAMtools and BEDtools (v2.29.0) (Quinlan and Hall, 2010). We used R to first smooth the read depth data which may show local high variation using the runmed function on 101 contiguous sites, then we kept only site showing a read depth of more than 5 to finally compute the ratio on Z-linked over Z pseudo-

autosomal read depth. We validated this approach by comparing molecular and *in silico* sexing of some F2s.

c. Linkage analysis.

We first reduced the VCF file using the SelectVariants module of GATK to include sites called in at least half of the samples. The QTL analysis was performed with R (v3.5.1) (R Core Team, 2018). We loaded the VCF file using vcfR (v1.10.0) (Knaus and Grünwald, 2017), filtered out sites with genotype quality (GQ) less than 30, a read depth (DP) less than 10 and with more than 20% of missing data. We retained alternatively fixed variants between parents (i.e., fully informative variants) and converted the genotype data into an R/qtl compatible format. We then run the scanone function of R/qtl (v1.46.2) (Broman et al., 2003) to map QTL in each cross or in the combined crosses using the expectation–maximization (EM) algorithm. We performed 1,000 permutations for each test.

Having identified QTLs, we selected a subset of around 400 markers total (number of markers per chromosome being proportional to the chromosome size) from the combined crosses and tested QTL interaction using the scantwo function from the same package with 1,000 permutations. We used a subset of markers because this function is computationally intensive. We tested the allelic dominance on the combined crosses using a custom R code.

d. Candidate gene prioritization.

To identify candidate genes involved in cercarial production, we first reduced the VCF file to sites within QTL regions identified previously. We loaded the VCF file in R using vcfR and refined the QTL regions by determining the chromosomal positions corresponding to a drop of 1.8 LOD from the highest LOD score of each QTL. We then retained only genotypes from the cross parents and filtered out sites with a GQ

lower than 30 and a DP lower than 6. We used a custom score (genoscore) to objectively prioritize candidate genes in each QTL region:

$$Genoscore = \frac{\sum_{i=1}^n (GSa_i \times GSb_i \times IS_i)}{l \times \sqrt{d_{peak}}} \times e \times 100$$

The genoscore is computed either by gene or gene's coding sequence (CDS). For each site i of a gene or CDS, we first computed a genotype score (GS) for each cross a and b by associating a score of 20 if genotypes were alternatively fixed in each cross parent or a score of 5 if fixed in one parent and unknown in the other (any other genotype combination received 0) and an impact score (IS) from the snpEff annotation (low impact or modifier mutation = 1, moderate impact = 5, high impact = 10). We summed the results for the n sites present in the gene or CDS. We weighed this sum using the length l of the gene or the CDS, and the square root of the distance d between the position of the highest LOD score (peak) and the start of the gene or the CDS. This was finally modulated by an expression factor e of 1 if the gene was expressed in mature sporocysts from shedding snails (Buddenborg et al., 2019) or in cercariae (Protasio et al., 2012) or -1 otherwise. Jupyter notebook and scripts used to obtain gene expression are available on Github (https://github.com/fdchevalier/Sm_gene_expression).

e. Gene annotations

To complete the gene annotation of the GFF file, we ran the HHsearch tool from the HH-suite (v3.3.0-0) (Steinegger et al., 2019) on each predicted protein sequences. This method relies on generating hidden Markov models (HMM) for a given sequence and compares HMM-HMM alignments. We compared our sequences to three databases (pdb70, scop70 and pfam) and selected the best match if the probability generated by HHsearch was at least of 50%. Gene identifications from the GFF file and the HHsearch analysis were combined in a table and used during the analysis of gene candidates. Jupyter notebook

and scripts used for this analysis are available on Github (<https://github.com/fdcchevalier/Sm-GFF-HHPred-table>).

Statistical analysis

All statistical analyzes and graphs were performed using R software (v3.5.1) (R Core Team, 2018). When data were not normally distributed (Shapiro test, $p < 0.05$), we compared results with non-parametric Kruskal-Wallis test followed by pairwise Wilcoxon-Mann-Whitney post-hoc test or a simple pairwise comparison Wilcoxon-Mann-Whitney test. When data followed a normal distribution, we used one-way ANOVA or a pairwise comparison Welch t -test. We performed correlation analysis using Pearson's correlation test. The confidence interval of significance was set to 95% and p -values less than 0.05 were considered significant.

ACKNOWLEDGEMENTS:

We thank Michael S. Blouin (Oregon State University) for the Bg26 snail line, Guillaume Mitta and Benjamin Gourbal (University of Perpignan France) for SmBRE *S. mansoni* population and Philip LoVerde (UT Health San Antonio) for the SmLE *S. mansoni* population. We thank the Vivarium of the Southwest National Primate Research Center (SNPRC) for providing rodent care. This research was supported by a Cowles fellowship (WL) from Texas Biomedical Research Institute (13-1328.021), and NIH R01AI133749 (TJCA) and was conducted in facilities constructed with support from Research Facilities Improvement Program grant C06 RR013556 from the National Center for Research Resources. SNPRC research at Texas Biomedical Research Institute is supported by grant P51 OD011133 from the Office of Research Infrastructure Programs, NIH.

WL, FDC and TJCA designed the experiments. WL, FDC, MMW, VM and GAA performed the experiments. WL and FDC performed the data analyses. WL and TJCA drafted the manuscript. All authors read and approved the final manuscript.

COMPETING INTERESTS

The authors declare that they have no competing interests.

REFERENCES:

- Acevedo MA, Dillemuth FP, Flick AJ, Faldyn MJ, Elder BD. 2019. Virulence-driven trade-offs in disease transmission: A meta-analysis. *Evolution; international journal of organic evolution* **73**:636–647. doi:10.1111/evo.13692
- Alizon S, Hurford A, Mideo N, Van Baalen M. 2009. Virulence evolution and the trade-off hypothesis: history, current state of affairs and the future. *J Evol Biol* **22**:245–259. doi:10.1111/j.1420-9101.2008.01658.x
- Alizon S, Michalakakis Y. 2015. Adaptive virulence evolution: the good old fitness-based approach. *Trends in ecology & evolution* **30**:248–254. doi:10.1016/j.tree.2015.02.009
- Anderson R.M. MRM. 1982. Coevolution of hosts and parasites. *Parasitology* **85**:411–26.
- Anderson TJC, LoVerde PT, Le Clec’h W, Chevalier FD. 2018. Genetic crosses and linkage mapping in schistosome parasites. *Trends Parasitol* **34**:982–996. doi:10.1016/j.pt.2018.08.001
- Behnke MS, Dubey JP, Sibley LD. 2016. Genetic mapping of pathogenesis determinants in toxoplasma gondii. *Annual review of microbiology* **70**:63–81. doi:10.1146/annurev-micro-091014-104353
- Behnke MS, Khan A, Lauron EJ, Jimah JR, Wang Q, Tolia NH, Sibley LD. 2015. Rhoptry proteins ROP5 and ROP18 are major murine virulence factors in genetically divergent south american strains of toxoplasma gondii. *PLoS genetics* **11**:e1005434. doi:10.1371/journal.pgen.1005434
- Behnke MS, Khan A, Wootton JC, Dubey JP, Tang K, Sibley LD. 2011. Virulence differences in Toxoplasma mediated by amplification of a family of polymorphic pseudokinases. *Proceedings of the National Academy of Sciences of the United States of America* **108**:9631–9636. doi:10.1073/pnas.1015338108
- Behnke MS, Zhang TP, Dubey JP, Sibley LD. 2014. Toxoplasma gondii merozoite gene expression analysis with comparison to the life cycle discloses a unique expression state during enteric development. *BMC genomics* **15**:350. doi:10.1186/1471-2164-15-350
- Berriman M, Haas BJ, LoVerde PT, Wilson RA, Dillon GP, Cerqueira GC, Mashiyama ST, Al-Lazikani B, Andrade LF, Ashton PD, Aslett MA, Bartholomeu DC, Blandin G, Caffrey CR, Coghlan A, Coulson R, Day TA, Delcher A, DeMarco R, Djikeng A, Eyre T, Gamble JA, Ghedin E, Gu Y, Hertz-Fowler C, Hirai H, Hirai Y, Houston R, Ivens A, Johnston DA, Lacerda D, Macedo CD, McVeigh P, Ning Z, Oliveira G, Overington JP, Parkhill J, Perteau M, Pierce RJ, Protasio AV, Quail MA, Rajandream M-A, Rogers J, Sajid M, Salzberg SL, Stanke M, Tivey AR, White O, Williams DL, Wortman J, Wu W, Zamanian M, Zerlotini A, Fraser-Liggett CM, Barrell BG, El-Sayed NM. 2009. The genome of the blood fluke *Schistosoma mansoni*. *Nature* **460**:352–358. doi:10.1038/nature08160
- Bonner KM, Bayne CJ, Larson MK, Blouin MS. 2012. Effects of Cu/Zn superoxide dismutase (sod1) genotype and genetic background on growth, reproduction and defense in *Biomphalaria glabrata*. *PLoS Negl Trop Dis* **6**:e1701. doi:10.1371/journal.pntd.0001701
- Bosch DE, Siderovski DP. 2013. G protein signaling in the parasite Entamoeba histolytica. *Experimental & molecular medicine* **45**:e15. doi:10.1038/emm.2013.30
- Broman KW, Wu H, Sen S, Churchill GA. 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics (Oxford, England)* **19**:889–890. doi:10.1093/bioinformatics/btg112
- Buddenborg SK, Kamel B, Hanelt B, Bu L, Zhang S-M, Mkoji GM, Loker ES. 2019. The in vivo transcriptome of Schistosoma mansoni in the prominent vector species Biomphalaria pfeifferi with supporting observations from Biomphalaria glabrata. *PLoS neglected tropical diseases* **13**:e0007013. doi:10.1371/journal.pntd.0007013
- Castle WE. 1921. AN IMPROVED METHOD OF ESTIMATING THE NUMBER OF GENETIC FACTORS CONCERNED IN CASES OF BLENDING INHERITANCE. *Science (New York, NY)* **54**:223. doi:10.1126/science.54.1393.223
- Chatel G, Fahrenkrog B. 2012. Dynamics and diverse functions of nuclear pore complex proteins. *Nucleus (Austin, Tex)* **3**:162–171. doi:10.4161/nucl.19674

Chevalier FD, Le Clec'h W, Alves de Mattos AC, LoVerde PT, Anderson TJC. 2016. Real-time PCR for sexing *Schistosoma mansoni* cercariae. *Mol Biochem Parasitol* **205**:35–38. doi:10.1016/j.molbiopara.2016.03.010

Chevalier FD, Valentim CLL, LoVerde PT, Anderson TJC. 2014. Efficient linkage mapping using exome capture and extreme QTL in schistosome parasites. *BMC Genomics* **15**:617. doi:10.1186/1471-2164-15-617

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 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

Cioli D, Pica-Mattoccia L, Moroni R. 1992. *Schistosoma mansoni*: hycanthone/oxamniquine resistance is controlled by a single autosomal recessive gene. *Experimental parasitology* **75**:425–432. doi:10.1016/0014-4894(92)90255-9

Cockerham CC. 1986. Modifications in estimating the number of genes for a quantitative character. *Genetics* **114**:659–664.

Collins, 3rd JJ, King RS, Cogswell A, Williams DL, Newmark PA. 2011. An atlas for *Schistosoma mansoni* organs and life-cycle stages using cell type-specific markers and confocal microscopy. *PLoS Negl Trop Dis* **5**:e1009. doi:10.1371/journal.pntd.0001009

Collins JNR, Collins JJ. 2017. Methods for studying the germline of the human parasite schistosoma mansoni. *Methods in molecular biology (Clifton, NJ)* **1463**:35–47. doi:10.1007/978-1-4939-4017-2_2

Criscione CD, van Paridon BJ, Gilleard JS, Goater CP. 2020. Clonemate cotransmission supports a role for kin selection in a puppeteer parasite. *Proceedings of the National Academy of Sciences of the United States of America* **117**:5970–5976. doi:10.1073/pnas.1922272117

Dahan-Pasternak N, Nasereddin A, Kolevzon N, Pe'er M, Wong W, Shinder V, Turnbull L, Whitchurch CB, Elbaum M, Gilberger TW, Yavin E, Baum J, Dzikowski R. 2013. PfSec13 is an unusual chromatin-associated nucleoporin of *Plasmodium falciparum* that is essential for parasite proliferation in human erythrocytes. *Journal of cell science* **126**:3055–3069. doi:10.1242/jcs.122119

Davies CM, Webster JP, Woolhous ME. 2001. Trade-offs in the evolution of virulence in an indirectly transmitted macroparasite. *Proc Biol Sci* **268**:251–257. doi:10.1098/rspb.2000.1367

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernysky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**:491–498. doi:10.1038/ng.806

Fentress SJ, Behnke MS, Dunay IR, Mashayekhi M, Rommereim LM, Fox BA, Bzik DJ, Taylor GA, Turk BE, Lichti CF, Townsend RR, Qiu W, Hui R, Beatty WL, Sibley LD. 2010. Phosphorylation of immunity-related GTPases by a *Toxoplasma gondii*-secreted kinase promotes macrophage survival and virulence. *Cell host & microbe* **8**:484–495. doi:10.1016/j.chom.2010.11.005

Files VS, Cram EB. 1949. A study on the comparative susceptibility of snail vectors to strains of *Schistosoma mansoni*. *The Journal of Parasitology* **35**:555–560.

Frank SA. 1996. Models of parasite virulence. *Q Rev Biol* **71**:37–78.

Gasser RB, Morahan G, Mitchell GF. 1991. Sexing single larval stages of *Schistosoma mansoni* by polymerase chain reaction. *Mol Biochem Parasitol* **47**:255–258. doi:10.1016/0166-6851(91)90187-B

Gérard C., Moné H. TA. 1993. *Schistosoma mansoni* – *Biomphalaria glabrata*: dynamics of the sporocyst population in relation to the miracidial dose and the host size. *Can J Zool* **71**:1880–1885.

Gower CM, Webster JP. 2004. Fitness of indirectly transmitted pathogens: restraint and constraint. *Evolution* **58**:1178–1184. doi:10.1111/j.0014-3820.2004.tb01698.x

Greenberg RM. 2013. New approaches for understanding mechanisms of drug resistance in schistosomes. *Parasitology* **140**:1534–1546. doi:10.1017/S0031182013000231

Hedrick PW. 1983. Genetics of populations, Series of books in biology. Science Books International.

- Houdusse A, Cohen C. 1996. Structure of the regulatory domain of scallop myosin at 2 Å resolution: implications for regulation. *Structure (London, England : 1993)* **4**:21–32. doi:10.1016/s0969-2126(96)00006-8
- Ittiprasert W, Mann VH, Karinshak SE, Coghlan A, Rinaldi G, Sankaranarayanan G, Chaidee A, Tanno T, Kumkhaek C, Prangtaworn P, Mentink-Kane MM, Cochran CJ, Driguez P, Holroyd N, Tracey A, Rodpai R, Everts B, Hokke CH, Hoffmann KF, Berriman M, Brindley PJ. 2019. Programmed genome editing of the omega-1 ribonuclease of the blood fluke, *J.avax.xml.bind.JAXBElement@36467c54*, *eLife* **8**. doi:10.7554/eLife.41337
- Jensen KH, Little TJ, Little T, Skorpung A, Ebert D. 2006. Empirical support for optimal virulence in a castrating parasite. *PLoS Biol* **4**:e197. doi:10.1371/journal.pbio.0040197
- Jones-Nelson O, Thiele EA, Minchella DJ. 2011. Transmission dynamics of two strains of *Schistosoma mansoni* utilizing novel intermediate and definitive hosts. *Parasitol Res* **109**:675–687. doi:10.1007/s00436-011-2299-2
- Kalbe M, Haberb B, Hertel J, Haas W. 2004. Heredity of specific host-finding behaviour in *Schistosoma mansoni* miracidia. *Parasitology* **128**:635–643. doi:10.1017/S0031182004005037
- Knaus BJ, Grünwald NJ. 2017. VCFR: a package to manipulate and visualize variant call format data in R. *Molecular ecology resources* **17**:44–53. doi:10.1111/1755-0998.12549
- Köster J, Rahmann S. 2018. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics (Oxford, England)* **34**:3600. doi:10.1093/bioinformatics/bty350
- Krautz-Peterson G, Bhardwaj R, Faghiri Z, Tararam CA, Skelly PJ. 2010. RNA interference in schistosomes: machinery and methodology. *Parasitology* **137**:485–495. doi:10.1017/S0031182009991168
- Lande R. 1981. The minimum number of genes contributing to quantitative variation between and within populations. *Genetics* **99**:541–553.
- Le Clec’h W, Anderson TJC, Chevalier FD. 2016. Characterization of hemolymph phenoloxidase activity in two *Biomphalaria* snail species and impact of *Schistosoma mansoni* infection. *Parasit Vectors* **9**:32. doi:10.1186/s13071-016-1319-6
- Le Clec’h W, Chevalier FD, McDew-White M, Allan F, Webster BL, Gouvras AN, Kinunghi S, Tchuem Tchuenté L-A, Garba A, Mohammed KA, Ame SM, Webster JP, Rollinson D, Emery AM, Anderson TJC. 2018. Whole genome amplification and exome sequencing of archived schistosome miracidia. *Parasitology* **1–9**. doi:10.1017/S0031182018000811
- Le Clec’h W, Diaz R, Chevalier FD, McDew-White M, Anderson TJC. 2019. Striking differences in virulence, transmission and sporocyst growth dynamics between two schistosome populations. *Parasites & vectors* **12**:485. doi:10.1186/s13071-019-3741-z
- Lewis FA, Stirewalt MA, Souza CP, Gazzinelli G. 1986. Large-scale laboratory maintenance of *Schistosoma mansoni*, with observations on three schistosome/snail host combinations. *J Parasitol* **72**:813–829.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**:1754–1760. doi:10.1093/bioinformatics/btp324
- Li Heng, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**:2078–2079. doi:10.1093/bioinformatics/btp352
- Li Hui, Wang B, Yang A, Lu R, Wang W, Zhou Y, Shi G, Kwon SW, Zhao Y, Jin Y. 2009. Ly-1 antibody reactive clone is an important nucleolar protein for control of self-renewal and differentiation in embryonic stem cells. *Stem cells (Dayton, Ohio)* **27**:1244–1254. doi:10.1002/stem.55
- Loker ES, Hofkin BV. 2015. Parasitology: A conceptual approach, 1st ed. Garland Science.
- Long S, Brown KM, Drewry LL, Anthony B, Phan IQH, Sibley LD. 2017. Calmodulin-like proteins localized to the conoid regulate motility and cell invasion by *Toxoplasma gondii*. *PLoS pathogens* **13**:e1006379. doi:10.1371/journal.ppat.1006379

- Mann VH, Suttiprapa S, Skinner DE, Brindley PJ, Rinaldi G. 2014. Pseudotyped murine leukemia virus for schistosome transgenesis: approaches, methods and perspectives. *Transgenic research* **23**:539–556. doi:10.1007/s11248-013-9779-3
- Marsolier J, Perichon M, DeBarry JD, Villoutreix BO, Chluba J, Lopez T, Garrido C, Zhou XZ, Lu KP, Fritsch L, Ait-Si-Ali S, Mhadhbi M, Medjkane S, Weitzman JB. 2015. Theileria parasites secrete a prolyl isomerase to maintain host leukocyte transformation. *Nature* **520**:378–382. doi:10.1038/nature14044
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 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
- Melman SD, Steinauer ML, Cunningham C, Kubatko LS, Mwangi IN, Wynn NB, Mutuku MW, Karanja DMS, Colley DG, Black CL, Secor WE, Mkoji GM, Loker ES. 2009. Reduced susceptibility to praziquantel among naturally occurring Kenyan isolates of Schistosoma mansoni. *PLoS neglected tropical diseases* **3**:e504. doi:10.1371/journal.pntd.0000504
- Messer PW, Ellner SP, Hairston NG. 2016. Can population genetics adapt to rapid evolution? *Trends in genetics : TIG* **32**:408–418. doi:10.1016/j.tig.2016.04.005
- Mitta G, Gourbal B, Grunau C, Knight M, Bridger JM, Théron A. 2017. The compatibility between biomphalaria glabrata snails and schistosoma mansoni: An increasingly complex puzzle. *Advances in parasitology* **97**:111–145. doi:10.1016/bs.apar.2016.08.006
- Mwangi IN, Sanchez MC, Mkoji GM, Agola LE, Runo SM, Cupit PM, Cunningham C. 2014. Praziquantel sensitivity of Kenyan Schistosoma mansoni isolates and the generation of a laboratory strain with reduced susceptibility to the drug. *International journal for parasitology Drugs and drug resistance* **4**:296–300. doi:10.1016/j.ijpddr.2014.09.006
- Otsuki H, Kaneko O, Thongkukiatkul A, Tachibana M, Iriko H, Takeo S, Tsuboi T, Torii M. 2009. Single amino acid substitution in Plasmodium yoelii erythrocyte ligand determines its localization and controls parasite virulence. *Proceedings of the National Academy of Sciences of the United States of America* **106**:7167–7172. doi:10.1073/pnas.0811313106
- Pattaradilokrat S, Culetton RL, Cheesman SJ, Carter R. 2009. Gene encoding erythrocyte binding ligand linked to blood stage multiplication rate phenotype in Plasmodium yoelii yoelii. *Proceedings of the National Academy of Sciences of the United States of America* **106**:7161–7166. doi:10.1073/pnas.0811430106
- Protasio AV, Tsai IJ, Babbage A, Nichol S, Hunt M, Aslett MA, De Silva N, Velarde GS, Anderson TJC, Clark RC, Davidson C, Dillon GP, Holroyd NE, LoVerde PT, Lloyd C, McQuillan J, Oliveira G, Otto TD, Parker-Manuel SJ, Quail MA, Wilson RA, Zerlotini A, Dunne DW, Berriman M. 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
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**:841–842. doi:10.1093/bioinformatics/btq033
- R Core Team. 2018. R: A language and environment for statistical computing. *R Foundation for Statistical Computing - Vienna, Austria*.
- Ramachandran S, Venugopal A, Kutty VR, Vinitha A, Divya G, Chitrasree V, Mullassari A, Pratapchandran NS, Santosh KR, Pillai MR, Kartha CC. 2014. Plasma level of cyclophilin A is increased in patients with type 2 diabetes mellitus and suggests presence of vascular disease. *Cardiovascular diabetology* **13**:38. doi:10.1186/1475-2840-13-38
- Randazzo PA, Andrade J, Miura K, Brown MT, Long YQ, Stauffer S, Roller P, Cooper JA. 2000. The Arf GTPase-activating protein ASAP1 regulates the actin cytoskeleton. *Proceedings of the National Academy of Sciences of the United States of America* **97**:4011–4016. doi:10.1073/pnas.070552297
- Reese ML, Zeiner GM, Saeij JPI, Boothroyd JC, Boyle JP. 2011. Polymorphic family of injected pseudokinases is paramount in Toxoplasma virulence. *Proceedings of the National Academy of Sciences of the United States of America* **108**:9625–9630. doi:10.1073/pnas.1015980108

- Rinaldi G, Eckert SE, Tsai IJ, Suttiprapa S, Kines KJ, Tort JF, Mann VH, Turner DJ, Berriman M, Brindley PJ. 2012. Germline transgenesis and insertional mutagenesis in *Schistosoma mansoni* mediated by murine leukemia virus. *PLoS pathogens* **8**:e1002820. doi:10.1371/journal.ppat.1002820
- Rollinson D, Stothard JR, Southgate VR. 2001. Interactions between intermediate snail hosts of the genus *Bulinus* and schistosomes of the *Schistosoma haematobium* group. *Parasitology* **123 Suppl**:S245–S260. doi:10.1017/s0031182001008046
- Saeij JPJ, Boyle JP, Collier S, Taylor S, Sibley LD, Brooke-Powell ET, Ajioka JW, Boothroyd JC. 2006. Polymorphic secreted kinases are key virulence factors in toxoplasmosis. *Science (New York, NY)* **314**:1780–1783. doi:10.1126/science.1133690
- Sahoo N, Labruyère E, Bhattacharya S, Sen P, Guillén N, Bhattacharya A. 2004. Calcium binding protein 1 of the protozoan parasite *Entamoeba histolytica* interacts with actin and is involved in cytoskeleton dynamics. *Journal of cell science* **117**:3625–3634. doi:10.1242/jcs.01198
- Sankaranarayanan G, Coghlan A, Driguez P, Lotkowska ME, Sanders M, Holroyd N, Tracey A, Berriman M, Rinaldi G. 2020. Large CRISPR-Cas-induced deletions in the oxamniquine resistance locus of the human parasite *S. mansoni*. *Wellcome open research* **5**:178. doi:10.12688/wellcomeopenres.16031.1
- Steinegger M, Meier M, Mirdita M, Vöhringer H, Haunsberger SJ, Söding J. 2019. HH-suite3 for fast remote homology detection and deep protein annotation. *BMC bioinformatics* **20**:473. doi:10.1186/s12859-019-3019-7
- Su L, Hersherberger RJ, Weissman IL. 1993. LYAR, a novel nucleolar protein with zinc finger DNA-binding motifs, is involved in cell growth regulation. *Genes & development* **7**:735–748. doi:10.1101/gad.7.5.735
- Tavalire HF, Blouin MS, Steinauer ML. 2016. Genotypic variation in host response to infection affects parasite reproductive rate. *Int J Parasitol* **46**:123–131. doi:10.1016/j.ijpara.2015.10.001
- Taylor S, Barragan A, Su C, Fux B, Fentress SJ, Tang K, Beatty WL, Hajj HE, Jerome M, Behnke MS, White M, Wootton JC, Sibley LD. 2006. A secreted serine-threonine kinase determines virulence in the eukaryotic pathogen *Toxoplasma gondii*. *Science (New York, NY)* **314**:1776–1780. doi:10.1126/science.1133643
- Théron A. 2015. Chronobiology of trematode cercarial emergence: from data recovery to epidemiological, ecological and evolutionary implications. *Advances in parasitology* **88**:123–164. doi:10.1016/bs.apar.2015.02.003
- Théron A, Combes C. 1988. Genetic analysis of cercarial emergence rhythms of *Schistosoma mansoni*. *Behavior genetics* **18**:201–209. doi:10.1007/BF01067842
- Théron A, Combes C. 1983. [Genetic analysis of the shedding pattern of *Schistosoma mansoni* cercariae by hybridization of species with early and late emission peaks]. *Comptes rendus des seances de l'Academie des sciences Serie III, Sciences de la vie* **297**:571–574.
- Theron A, Rognon A, Gourbal B, Mitta G. 2014. Multi-parasite host susceptibility and multi-host parasite infectivity: A new approach of the *Biomphalaria glabrata*/*Schistosoma mansoni* compatibility polymorphism. *Infect Genet Evol* **26**:80–88. doi:10.1016/j.meegid.2014.04.025
- Thomas VP, Williamson VM. 2013. Segregation and mapping in the root-knot nematode *Meloidogyne hapla* of quantitatively inherited traits affecting parasitism. *Phytopathology* **103**:935–940. doi:10.1094/PHYTO-03-13-0074-R
- Tuo W, Fetterer R, Jenkins M, Dubey JP. 2005. Identification and characterization of *Neospora caninum* cyclophilin that elicits gamma interferon production. *Infection and immunity* **73**:5093–5100. doi:10.1128/IAI.73.8.5093-5100.2005
- Ünal CM, Steinert M. 2014. Microbial peptidyl-prolyl cis/trans isomerases (PPIases): virulence factors and potential alternative drug targets. *Microbiology and molecular biology reviews : MMBR* **78**:544–571. doi:10.1128/MMBR.00015-14
- Valentim CLL, Cioli D, Chevalier FD, Cao X, Taylor AB, Holloway SP, Pica-Mattoccia L, Guidi A, Basso A, Tsai IJ, Berriman M, Carvalho-Queiroz C, Almeida M, Aguilar H, Frantz DE, Hart PJ, LoVerde PT, Anderson TJC.

2013. Genetic and molecular basis of drug resistance and species-specific drug action in Schistosome parasites. *Science* **342**:1385–1389. doi:10.1126/science.1243106
- Vasudevan A, Kumar V, Chiang YN, Yew JY, Cheemadan S, Vyas A. 2015. α 2u-globulins mediate manipulation of host attractiveness in *Toxoplasma gondii*-*Rattus norvegicus* association. *The ISME journal* **9**:2112–2115. doi:10.1038/ismej.2015.33
- Verjee MA. 2019. Schistosomiasis: Still a cause of significant morbidity and mortality. *Research and reports in tropical medicine* **10**:153–163. doi:10.2147/RRTM.S204345
- Wang B, Collins JJ, Newmark PA. 2013. Functional genomic characterization of neoblast-like stem cells in larval *Schistosoma mansoni*. *eLife* **2**:e00768. doi:10.7554/eLife.00768
- Wang B, Lee J, Li P, Saberi A, Yang H, Liu C, Zhao M, Newmark PA. 2018. Stem cell heterogeneity drives the parasitic life cycle of *Schistosoma mansoni*. *eLife* **7**. doi:10.7554/eLife.35449
- Wang J, Paz C, Padalino G, Coghlan A, Lu Z, Gradinaru I, Collins JNR, Berriman M, Hoffmann KF, Collins JJ. 2020. Large-scale RNAi screening uncovers therapeutic targets in the parasite , *Science (New York, NY)* **369**:1649–1653. doi:10.1126/science.abb7699
- Wang P, Heitman J. 2005. The cyclophilins. *Genome biology* **6**:226. doi:10.1186/gb-2005-6-7-226
- Webster JP, Gower CM, Blair L. 2004. Do hosts and parasites coevolve? Empirical support from the *Schistosoma* system. *Am Nat* **164 Suppl 5**:S33–S51. doi:10.1086/424607
- Wei Y, Jinchuan Y, Yi L, Jun W, Zhongqun W, Cuiping W. 2013. Antiapoptotic and proapoptotic signaling of cyclophilin A in endothelial cells. *Inflammation* **36**:567–572. doi:10.1007/s10753-012-9578-7
- Wendt G, Zhao L, Chen R, Liu C, O'Donoghue AJ, Caffrey CR, Reese ML, Collins JJ. 2020. A single-cell RNA-seq atlas of , *Science (New York, NY)* **369**:1644–1649. doi:10.1126/science.abb7709
- Wendt GR, Collins JJ. 2016. Schistosomiasis as a disease of stem cells. *Current opinion in genetics & development* **40**:95–102. doi:10.1016/j.gde.2016.06.010
- World Health Organization. 2019. Schistosomiasis. <https://www.who.int/news-room/fact-sheets/detail/schistosomiasis>
- Wright SW, Morton NE. 1968. Genetic studies on cystic fibrosis in Hawaii. *American journal of human genetics* **20**:157–169.
- Yamamoto M, Ma JS, Mueller C, Kamiyama N, Saiga H, Kubo E, Kimura T, Okamoto T, Okuyama M, Kayama H, Nagamune K, Takashima S, Matsuura Y, Soldati-Favre D, Takeda K. 2011. ATF6beta is a host cellular target of the *Toxoplasma gondii* virulence factor ROP18. *The Journal of experimental medicine* **208**:1533–1546. doi:10.1084/jem.20101660
- Yau W-L, Blisnick T, Taly J-F, Helmer-Citterich M, Schiene-Fischer C, Leclercq O, Li J, Schmidt-Arras D, Morales MA, Notredame C, Romo D, Bastin P, Späth GF. 2010. Cyclosporin A treatment of *Leishmania donovani* reveals stage-specific functions of cyclophilins in parasite proliferation and viability. *PLoS neglected tropical diseases* **4**:e729. doi:10.1371/journal.pntd.0000729
- You H, Mayer JU, Johnston RL, Sivakumaran H, Ranasinghe S, Rivera V, Kondrashova O, Koufariotis LT, Du X, Driguez P, French JD, Waddell N, Duke MG, Ittiprasert W, Mann VH, Brindley PJ, Jones MK, McManus DP. 2021. CRISPR/Cas9-mediated genome editing of *Schistosoma mansoni* acetylcholinesterase. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **35**:e21205. doi:10.1096/fj.202001745RR

857 **SUPPLEMENTARY FIGURES:**

858

859 **Supplementary figure 1: Impact of *S. mansoni* gender on the cercarial production for F1s (A-B) and**
860 **F2s (C-D). (A)** Male sporocysts produced significantly less cercariae than female sporocysts in F1A cross
861 parasite but **(B)** there is no difference driven by the gender of the parasites for the F1B cross. For F2
862 progeny, gender has no impact on cercarial production in cross A **(C)**, but male sporocysts produced
863 significantly less cercariae than female sporocysts in cross B **(D)**. NS: No significant difference in cercarial
864 production between the two considered groups; * $p < 0.05$; ** $p \leq 0.02$; *** $p \leq 0.002$.

865

867 **Supplementary Figure 2. Influence of the shedding week on the genetic architecture of cercarial**
868 **production. (A)** On shedding week 1, none of the QTLs identified pass the permutation threshold **(B)**
869 On shedding week 2, the only QTL involved in the transmission trait is located on chr. 5. **(C)** On
870 shedding week 3, QTL on chr. 3 is the only one with a significant LOD score. **(D)** Finally on shedding
871 week 4, QTL on chr. 1 seems to be the one explaining most of the phenotype, along with QTL on chr. 2
872 and on chr. 3. Red line refers to assembled chromosome and black line to unassembled scaffolds
873 assigned to chromosome or still unassigned. The blue dotted line corresponds to the 1,000
874 permutation threshold.

875

876

877 **Supplementary Figure 3. Correlation between laccase-like activity and hemoglobin rate measured in**
878 **the hemolymph of infected intermediate snail host *Biomphalaria glabrata* and the cercarial**
879 **production for F1 (A-C) and F2 (D-F) *S. mansoni* parasite progeny (7.5 weeks post-infection).** We did
880 not observe significant correlations neither **(A)** between laccase-like activity and the average
881 production of cercariae, nor **(B)** between the hemoglobin rate and the average cercarial production in
882 F1 *S. mansoni* progeny. This result was expected as all the individual exhibited heterozygous genotype
883 (i.e. LH) and showed an intermediate phenotype in their transmission stage production compared to

884 their respective parents. However, we observed a significant negative correlation **(D)** between the
 885 average cercariae produced by F2 parasite progeny and the total laccase-like activity in the hemolymph
 886 of the infected snail (Pearson's test, coef. = -0.56, $p < 0.0001$) and **(E)** between these cercarial
 887 production and the hemoglobin rate measured in infected snail hemolymph (Pearson's test, coef. = -
 888 0.50, $p < 0.0001$). For both snail physiological parameters measured in the hemolymph, we highlighted
 889 a strong positive correlation (For F1: Pearson's test, coef. = 0.72, $p < 0.0001$ and for F2: Pearson's test,
 890 coef. = 0.66, $p < 0.0001$) whatever the parasite generation analyzed.

Supplementary Figure 4. Example of read depth across the Z chromosome in males and females. The left column represents male samples and the right column female samples. The top rows are known sex samples (parents of cross A), the two other rows are samples for which we predicted sex *in silico*. The grey dots represent the smoothed read depth, the red line the local regression of the read depth obtained from the loess function.

898 **Supplementary Table 1. Complete list of candidate genes linked to transmission stage production in**
899 ***S. mansoni* parasite.** This table gives summary information for each gene under the QTL regions and
900 detailed information about the variants in these regions.