# Genomic analysis of a parasite invasion: colonization of the Americas by the blood fluke, *Schistosoma mansoni*

Roy N. Platt II*[1], Winka Le Clec'h*[1], Frédéric D. Chevalier*[1], Marina McDew-White[1], Philip T. LoVerde[2], Rafael R. de Assis[3], Guilherme Oliveira[3,4], Safari Kinunghi[5], Amadou Garba Djirmay[6], Michelle L. Steinauer[7], Anouk Gouvras[8], Muriel Rabone[8], Fiona Allan[9,10], Bonnie L. Webster[8,10], Joanne P. Webster[9,10], Aidan Emery[8,10], David Rollinson[8,10], Timothy J. C. Anderson[1]

[1] *Texas Biomedical Research Institute, San Antonio, San Antonio, TX, United States*

[2] *University of Texas Health Science Center, San Antonio, San Antonio, TX, United States*

[3] *Centro de Pesquisas René Rachou - Fiocruz/MG, Belo Horizonte, Brazil*

[4] *Instituto Tecnológico Vale, Belém, Brazil*

[5] *National Institute for Medical Research, Mwanza, Tanzania.*

[6] *Réseau International Schistosomiases Environnemental Aménagement et Lutte (RISEAL), Niamey, Niger.*

[7] *Western University of Heath Sciences, Corvallis, OR, United States*

[8] *Natural History Museum, London, United Kingdom*

[9] *Centre for Emerging, Endemic and Exotic Diseases, Department of Pathobiology and Population Sciences, Royal Veterinary College, University of London, Hertfordshire, United Kingdom*

[10] *London Centre for Neglected Tropical Disease Research, Imperial College London, School of Public Health, London, United Kingdom*

## Abstract

*Schistosoma mansoni,* a snail-vectored, blood fluke that infects humans, was introduced into the Americas from Africa during the Trans-Atlantic slave trade. As this parasite shows strong specificity to the snail intermediate host, we expected that adaptation to S. American *Biomphalaria* spp. snails would result in population bottlenecks and strong signatures of selection. We scored 475,081 single nucleotide variants (SNVs) in 143 *S. mansoni* from the Americas (Brazil, Guadeloupe, and Puerto Rico) and Africa (Cameroon, Niger, Senegal, Tanzania, and Uganda), and used these data to ask: (i) Was there a population bottleneck during colonization? (ii) Can we identify signatures of selection associated with colonization? And (iii) what were the source populations for colonizing parasites? We found a 2.4-2.9-fold reduction in diversity and much slower decay in linkage disequilibrium (LD) in parasites from East to West Africa. However, we observed similar nuclear diversity and LD in West Africa and Brazil, suggesting no strong

bottlenecks and limited barriers to colonization. We identified five genome regions showing selection in the Americas, compared with three in West Africa and none in East Africa, which we speculate may reflect adaptation during colonization. Finally, we infer that unsampled African populations from central African regions between Benin and Angola, with contributions from Niger, are likely the major source(s) for Brazilian *S. mansoni*. The absence of a bottleneck suggests that this is a rare case of a serendipitous invasion, where *S. mansoni* parasites were preadapted to the Americas and were able to establish with relative ease.

**Keywords**

co-dispersal, migration, human parasite, exome, Africa, Brazil

## Introduction

Genomic characterization of parasites and pathogens is increasingly used as an aid to traditional epidemiological methods in reconstructing transmission patterns (de Oliveira et al., 2020; Nadeau, Vaughan, Scire, Huisman, & Stadler, 2021). On a longer time-scale genomic data can be used to understand biological invasions of pathogens into new continents, just as these methods are used for investigating biological invasions in free-living organisms (Rius, Bourne, Hornsby, & Chapman, 2015; Sherpa & Després, 2021). Such methods can determine the colonization route, source population, number of colonization events, whether diversity is reduced during colonization and evidence for adaptation in colonizing populations. Examining the consequences of historic invasions can inform our understanding of extant invasions.

The Trans-Atlantic slave trade lasted from 1502-1888 when the last remaining slave ports in Brazil were shutdown (Bergad, 2007). During this time, more than 12 million people were trafficked from Africa to slave ports in the Americas, representing one of the largest forced migration events in human history (Eltis, 2001). Along with the human cargo, a number of human pathogens were introduced into the Americas as well. For example, Parvo and hepatitis B viruses were successfully introduced into the Americas and rapidly spread through native populations leading to large scale outbreaks (Guzmán-Solís et al., 2021). *Trypansoma brucei*, the etiological agent for African Sleeping Sickness, was also introduced but failed to establish due to the absence of its' intermediate host; the Tsetse fly (Steverding, 2020). Today viable populations of pathogens including Herpes simplex virus 2 (Forni et al., 2020), yellow fever virus (Bryant, Holmes, & Barrett, 2007), the parasitic nematode *Wuchereria bancrofti* (Small et al., 2019), among others (Steverding, 2020) are all a direct result of introductions during the Trans-Atlantic slave trade. In some cases, the genetic signatures of the introduction are still visible. For

example, genetic diversity in South American *Leishmania chagasi* populations is halved and the effective population size (*Ne*) is reduced from 43.6M to 15.5K when compared to source populations in Africa (Leblois, Kuhls, François, Schönian, & Wirth, 2011; Schwabl et al., 2021). Here, we focus on successful invasion of the human-parasitic trematode, *Schistosoma mansoni*.

*S. mansoni* is distributed from Oman, through sub-Saharan Africa, to the Caribbean and countries along the eastern coast of South America. Phylogenetic evidence indicates that *S. mansoni* in West Africa and the Americas are closely related (Crellen et al., 2016; Desprès, Imbert-Establet, & Monnerot, 1993; Fletcher, LoVerde, & Woodruff, 1981; Morgan et al., 2005; Webster et al., 2013) and these observations, along with demographic reconstructions (Crellen et al., 2016), indicate a recent origin of *S. mansoni* in the Americas. As a result, there is strong evidence that *S. mansoni* co-migrated to the Americas during the forced human migrations of the Trans-Atlantic slave trade (Files, 1951). Furthermore, reduced diversity in mitochondrial haplotypes (Desprès et al., 1993; Fletcher et al., 1981; Morgan et al., 2005; Webster et al., 2013) in South American *S. mansoni* suggests the presence of a bottleneck during parasite establishment.

Our central goal was to use parasite genomic data to investigate this human-mediated, biological invasion and the impacts of a relatively recent, trans-continental, migration event. Parasites in the genus *Schistosoma* have a complex life cycle involving human definitive hosts and snail intermediate hosts (reviewed in Anderson & Enabulele, 2021). Eggs are expelled in human feces (*S. man*soni and *S. japonicum*) or urine (*S. haematobium*). Larvae (miracidia) hatch in fresh water and infect receptive snails. Once inside the snail host, the schistosomes reproduce asexually, and 2$^{nd}$ stage larvae (cercariae) are released back into the water where they infect humans, mature into adult worms, and restart their life cycle. *S. mansoni* is diploid, with a well characterized 363Mb genome (Berriman et al., 2009; International Helminth Genomes Consortium, 2019; Protasio et al., 2012), ZW sex determination, obligate sexual reproduction of adult worms, and a relatively long life-span; from 5-10 years (Fulford, Butterworth, Ouma, & Sturrock, 1995).

The distribution of the intermediate snail host is a major driver of schistosome distribution. *S. haematobium*, a sister-taxon to *S. mansoni*, infects a different snail host; *Bulinus* spp. While people infected with both *S. mansoni* and *S. haematobium* were transported to S. America, only *S. mansoni* would have been able to establish due to the presence of the *Biomphalaria* spp. intermediate snail host(s) prior to their arrival (Morgan et al., 2005). *S. mansoni* shows strong specificity for species, and even strains of snails in the genus *Biomphalaria* (Webster & Woolhouse, 1998), however the *Biomphalaria* species assemblages differ between the Americas and Africa (DeJong et al., 2001). *B. pfeifferi*, *B. sudanica*, and *B. alexandrina*, are the primary intermediate hosts in Africa (DeJong et al., 2001), while

*B. glabrata*, *B. tenagophila*, *B. straminea*, are the known snail hosts in South America (Vidigal et al., 2000). *S. mansoni* infections can impact the reproductive viability of their snail hosts, and there are strong co-evolutionary interactions driving resistance to infection in snails and for infectivity in parasites (Davies, Webster, & Woolhous, 2001; Theron, Rognon, Gourbal, & Mitta, 2014; Webster, Gower, & Blair, 2004). Several schistsome resistance genes have been localized within the snail genome (Tennessen et al., 2020; Tennessen et al., 2015) and polymorphic loci in both snail and parasites are thought to determine compatibility between snail and parasite (Mitta et al., 2017; Webster & Woolhouse, 1998; Woolhouse & Webster, 2000). Based on these observations, we hypothesize that the adaptation to novel, *Biomphalaria* spp. hosts would place strong selective pressures on *S. mansoni* as it became established in the Americas.

Adult schistosomes live in the blood vessels, making them difficult to sample. Genome and exome sequencing of schistosomes is now possible using whole genome amplification of miracidia larvae isolated from feces or urine (Doyle et al., 2019; Le Clec'h et al., 2018; Shortt et al., 2017) and several genome scale population analyses have recently been published (Berger et al., 2021; Platt et al., 2019; Shortt et al., 2017). Our goal is to address the following questions with the available sequence data from both the Africa (Niger, Senegal, Uganda, and Tanzania) and the Americas (Caribbean, Brazil): (i) Are the genomic data consistent with a West African origin of colonizing schistosome populations?; (ii) is there evidence for genetic bottlenecks during colonization; (iii) are there genomic signatures suggesting adaptation of colonizing parasites to the Americas; (iv) can we determine the source country or countries for Americas parasite populations?

## Materials and Methods

*Data and sample information* – We examined published exomic and genomic data from 178 individual *Schistosoma* samples/isolates, from multiple geographical locations, available from three studies (Berriman et al., 2009; Chevalier et al., 2019; Crellen et al., 2016). All exome data is from Chevalier *et al*. (2016) and Chevalier *et al*. (2019). These data were generated from individual larval miracidia hatched from *S. mansoni* eggs and preserved on FTA cards. Exome libraries were generated via whole genome amplification followed by targeted capture of the exome (Le Clec'h et al., 2018). This method specifically targets 95% (14.81 Mb) of the exome with 2x tiled probes. The whole genome sequence data came from adult worms cultured through laboratory rodents and snails for two or more generations before whole genome library prep and sequencing (Berriman et al., 2009; Crellen et al., 2016; International Helminth Genomes Consortium, 2019). Sample origins are shown in Figure 2. Detailed metadata is available for

each sample in Supplemental Table 1 including country of origin, species identification, NCBI Short Read Archive (SRA) accession, *etc*.

*Computational environment* –We used conda v4.8.3 to manage virtual environments for all analyses. Sequence read filtering through genotyping steps were documented in a Snakemake v5.18.1 (Köster & Rahmann, 2012) workflow and all other analyses were performed in a series of Jupyter v1.0.0 notebooks. The code for this project including shell scripts, Snakemake workflows, notebooks, and environmental yaml files are available at https://github.com/nealplatt/sch_man_nwinvasion/releases/tag/v0.2 (last accessed 21 Oct 2021) and accessioned at https://doi.org/10.5281/zenodo.5590460 (last accessed 21 Oct 2021).

*Genotyping* – Paired end reads were quality filtered with trimmomatic v0.39 (Bolger, Lohse, & Usadel, 2014) with the following parameters 'LEADING:10 TRAILING:10 SLIDINGWINDOW:4:15 MINLEN:36'. Filtered reads were mapped to the *S. mansoni* genome (GenBank Assembly accession: GCA_000237925.3) using bwa v0.7.17-r1188 (Li & Durbin, 2010). We allowed up to 15 mismatches per 100 bp read (-n 15) to account for divergence between *S. mansoni* and *S. rodhaini*. Mapped single and paired reads were merged into a single file and all optical/PCR duplicates were removed with GATK v4.1.2.0's (McKenna et al., 2010) MarkDuplicates. Single nucleotide variants (SNVs)were called with HaplotypeCaller and GenotypeGVCFs on a contig-by-contig basis, combined into a gvcf per individual, and finally merged into a single gvcf for the entire dataset. We used a high quality SNV data from Le Clec'h *et al.* (2021) as a training dataset for variant recalibration and scored SNV quality using '-an SOR -an MQ -an MQRankSum -an ReadPosRankSum'. Sensitivity Tranches (--truth-sensitivity-tranche) were set at 100, 99.5, 99, 97.5, 95, and 90. We re-calibrated SNVs using the 97.5 sensitivity tranche and filtered low confidence sites with the following set of filters (--filter-expression) QD < 2.0, MQ < 30.0, FS > 60.0, SOR > 3.0, MQRankSum < -12.5, and ReadPosRankSum < -8.0". All genotyping steps from read filtering through variant re-calibration were contained within a single Snakemake v5.18.1 (Köster & Rahmann, 2018) script.

We used VCFtools v0.1.16 (Danecek et al., 2011) for additional rounds of filtering. First, we removed low quality sites with quality score <25, read depth <12, and non-biallelic sites. Second, we removed sites and individuals with a genotyping rate less than 50%. Third, we removed all sites that were on unresolved haplotigs by retaining only those SNVs that were on one of seven autosomal scaffolds (GenBank Nucleotide accessions: HE601624.2-30.2), the sex-linked ZW scaffold (HE601631.2), or the mitochondria (HE601612.2). Finally, for analyses requiring unlinked SNVs, we filtered linked sites

within 250Kb windows using Plink v1.90b4 (Purcell et al., 2007) with the following parameters "—indep-pairwise 250kb 1 0.20".

*Summary Statistics* – We quantified read depth per probed-exome region with MosDepth v0.2.5 (Pedersen & Quinlan, 2018) and calculated genome-wide summary statistics for each population, including $F_3$, $F_{ST}$, Tajima's *D*, π, and the Watterson estimator (Θ) with scikit-allele v1.2.1 (Miles, Ralph, Rae, & Pisupati, 2019). We examined genome regions that were targeted by the Le Clec'h *et al*. (2018) probe set for these calculations, non-target regions (i.e. non-exomic) were ignored since most samples lacked information from these regions. $F_{ST}$ between populations was calculated form the average Weir-Cockerham $F_{ST}$ (Weir & Cockerham, 1984) in windows of 100 SNVs. Effective population size (*Ne*) was estimated from Θ and the mutation rate (μ = 8.1 e-9 per base per generation; Crellen et al., 2016) with:

$$Ne = \frac{\Theta}{4\mu}$$

We examined linkage disequilibrium (LD) within each population by calculating $r^2$ (--r2) values with PLINK v1.90b6.18 (Purcell et al., 2007). We excluded invariant sites from the analyses. Intra-autosomal, pairwise comparisons between SNVs within 1Mb of one another were allowed by setting the following parameters: "--ldwindow 1000000", "--ld-window-kb 1000", and "--ld-window-r2 to 0.0". $r^2$ values were then binned into 500 bp windows and averaged for each population using the R v3.6.1 stats.bin function in the fields v11.6 (Nychka, Furrer, Paige, & Sain, 2017) library. We used local regression to smooth the binned $r^2$ values with the loessMod function in the base R v3.6.1 package and a span size of 0.5.

We used a Pearson Mantel test to examine correlation between genetic and physical distance. Since we did not have exact collection coordinates from whole genome samples, or they were lab derived, we excluded them from the analyses and instead focused only on the *S. mansoni* exome samples. We calculated pairwise p-distances with VCF2Dis (https://github.com/BGI-shenzhen/VCF2Dis; commit: b7684d3, accessed 13 Feb 2021) and physical distances between samples the Python haversine 2.3.0 module. Finally, we used the mantel() function in the scikit-bio 0.2.1 Python library to conduct a Pearson Mantel test that included 1,000 permutations.

*Population structure and admixture* - We examined population substructure using PCA and Admixture with unlinked autosomal SNVs (described above). Two PCAs were calculated in PLINK v1.90b6.18, with and without the *S. rodhaini* samples. Population ancestry was estimated with ADMIXTURE v1.3.0

(Alexander, Novembre, & Lange, 2009). We examined between $k$=1 to $k$=20 populations and used the Cross-validation scores and the Evanno *et al*. (2005) method were used to determine a range of viable $k$'s. Q estimates were used as proxy for ancestry fractions.

We examined D (Patterson et al., 2012), $D_3$ (Hahn & Hibbins, 2019) and $F_3$ (Patterson et al., 2012) to identify gene flow between *S. rodhaini* and *S. mansoni* populations with emphasis on the Tanzanian *S. mansoni* population since it is from East Africa as are the *S. rodhaini* samples. For $D_3$, we calculated the mean-pairwise (Euclidean) distances between populations using scikit-allele's allel.pairwise_distance() function. To determine significance, we used 1,000 block bootstrap replicates of 1,000 SNV blocks. We calculated the average $F_3$ across the genome in blocks of 100 variants. Here we ran multiple test that included some combination of an African *S. mansoni* population (Niger, Senegal, and Tanzania) as the test group and Brazilian *S. mansoni* and *S. rodhaini* as the potential source populations. D, or the ABBA-BABBA statistic, averaged over blocks of 1,000 variants assuming a phylogeny of (((*a*, Tanzania), *S. rodhaini*), *S. margrebowiei*), where the *a* population was either Brazil, Niger, or Senegal. D, $D_3$ and $F_3$ values were calculated using scikit-allel.

*Phylogenetics* – We used three different phylogenetic methods to visualize relationships among sampled schistosomes: a mitochondrial haplotype network, a coalescent-based species tree, and a phylogenetic network.

(i) Mitochondrial haplotype network. We extracted mitochondrial SNVs from all *S. mansoni* individuals with VCFtools and converted the subsequent VCF file to Nexus format with vcf2phylip v2.0 (Ortiz, 2019). A median joining network (ε = 0) was created in from the mitochondrial haplotypes with PopArt v1.7 (Leigh & Bryant, 2015).

(ii) Coalescent-based species tree. We generated a coalescent-based species tree with SVDQuartets (Chifman & Kubatko, 2015) packaged in PAUP* v4.0.a.build166 (Swofford, 2003). We examined parsimony-informative, autosomal SNVs by removing private alleles (singleton and doubletons). All samples were assigned to a population based on their country of origin (ex. Niger, Puerto Rico, Brazil, Cameroon, etc.) except for the lab-derived, Caribbean samples. Each of these samples was considered to represent an individual population given their histories of extensive lab passage (ex. Guadeloupe1, Guadeloupe2, Puerto Rico). We randomly evaluated 100,000 random quartets and bootstrapped the quartet tree with 1,000 standard replicates. The tree was rooted on the single *S. margrebowiei* individual.

(iii) Phylogenetic network. We used a phylogenetic network to visualize and quantify migration among schistosome populations. We only included *S. mansoni* populations with more than 4 individuals, which excluded all whole genome samples from this analysis including those from the Caribbean and the *S. rodhaini* samples. We used autosomal SNVs after filtering linked sites in 250Kb blocks with PLINK v1.90b6.18 and then used TreeMix v1.12 (Pickrell & Pritchard, 2012) to generate the phylogenetic network. This analysis used a co-variance matrix generated from blocks of 500 SNVs without sample-size correction ("--noss") and the number of migration events was limited to 3.

*Selection* –We scanned the genome to identify regions under selection using haplotype (H-scan v1.3; Schlamp et al., 2016), allele-frequency (SweepFinder2 v2.1; DeGiorgio, Huber, Hubisz, Hellmann, & Nielsen, 2016), and PCA-based (pcadapt v4.3.3; Luu, Bazin, & Blum, 2017) methods. In addition, to avoid false positives, we used msprime v0.7.4 (Kelleher, Etheridge, & McVean, 2016) to conduct simulations to estimate the range of values expected under neutrality from H-Scan and SweepFider2.

(i) Neutral simulations. We used msprime v0.7.4 (Kelleher et al., 2016) to simulate a set of neutrally evolving SNVs along a single chromosome for each population and then used the simulated data with H-Scan and SweepFinder2 to define the range of values expected in the absence of selection. For these simulations we used a mutation rate ($\mu$ = 8.1 e-9 per base per generation; Crellen et al., 2016) and recombination rate (3.4e-8 per base per generation; Criscione, Valentim, Hirai, LoVerde, & Anderson, 2009) from previous work on *S. mansoni*. Population-specific estimates of *Ne* are described above. The chromosome length was set to 88.9 Mb which is equal to chromosome 1 (HE601624.2) in the *S. mansoni* assembly. The number of chromosomes sampled per msprime run was equal to the number of samples we collected in each population. We used these parameters to perform 342 simulations for each population, roughly equivalent to 100 genomes worth of simulated data. Next, we down sampled the SNVs along the entire simulated chromosome so that they were comparable with our targeted sequencing approach (i.e. only SNVs from "exomic" regions). Since the simulated chromosomes were the same size as chromosome 1 we transposed the chr1 annotation onto the simulated chromosome and extracting only those SNVs occurring in regions accessible by our biotinylated probes with VCFtools. The simulated data was run through H-Scan and SweepFinder2 in parallel with the actual SNV data to establish a range of neutral values for each test (described above).

(ii) H-scan. This method measures the length of homozygous haplotypes to identify regions under selection. Strong selection on adaptive alleles drives SNVs under selection and any linked alleles to high frequency and reduces homozygosity in the surrounding region. For each population, we converted autosomal SNVs from VCF to H-Scan format using vcf2hscan.py script from vcf2phylip v2.0 (Ortiz, 2019) and ran H-Scan on each chromosome with a maximum gap length (-g) of 10Kb. *H* values were smoothed ($H_{smoothed}$) by median filtering values in 201 SNV windows (step size = 1) using the medfilt function in the SciPy v1.5.2 (Virtanen et al., 2020) for visualization purposes.

(iii) SweepFinder2. This method uses deviations in allele frequency from a neutral expectation to estimate the selection while accounting for the possibility of background selection via a likelihood ratio (LR) test (DeGiorgio et al., 2016). Empirical site-frequency spectra were calculated for each population and within each population LR was estimated along each autosome individually. We examined grid points ('g'), or window sizes, of 1, 5, 10, and 20 Kb with minimal impact on the results. Downstream analyses are reported on the runs with 'g' = 1kb.

(iv) *pcadapt* – We used the R v4.0.5 package pcadapt v4.3.3 (Luu et al., 2017) to identify highly differentiated loci among populations via variants associated with population structure as identified by PCA. We only included all samples from Brazil, Niger, and Senegal since our primary goal was to identify variants involved in adaptation the Americas. Rare variants (MAF<5%) were excluded with VCFtools v0.1.16. We identified the appropriate number of principal components from the data by running an initial *pcadapt* run with 20 populations (K=20) and LD filtered variants (LD.clumping = list(size = 100, thr = 0.2)). The major break in the subsequent scree plot was used as the optimal K choice. We used a second pcadapt run with the optimal K and the same LD filtering parameters as the initial run to assign *p* values to each site. Finally, we adjusted *p* values for multiple tests with Bonferroni correction and an $\alpha = 0.05$ to identify SNV outliers associated with population differentiation.

(v) Identifying regions of selection- We identified regions potentially under positive selection using a three-step process. First, we identified SNVs whose H-Scan and Sweepfinder2 values were in the 99$^{th}$ percentile of and greater than the neutral thresholds established with msprime. These were SNVs with the strongest signal of selection. Then, we expanded from the SNV to a broader region by merging all variants within 333,333 bp whose H-Scan or Sweepfinder2 values were greater than the neutral thresholds. Finally, we looked for pcadapt outliers in each region. These regions are referred to as "putative selected regions" or "putative regions of selection". Gene

294 names and putative functions were taken from UniProtKB (release 2020_06) or HHsearch
295 (Steinegger et al., 2019) annotations from Le Clec'h et al. (2021). The entire process is
296 summarized in Figure 1.
297
298 **Results**
299 *Summary of Sequence Data* – After genotyping and filtering, we removed 25 of the 178 samples with
300 low numbers of reads, poor coverage, or low genotyping rates. The final dataset included 135 *S. mansoni*
301 (exome), 8 *S. mansoni* (genome), 8 *S. rodhaini* (exome), 1 *S. rodhaini* (genome), and 1 *S. margrebowiei*
302 (genome). We genotyped 1,823,890 sites, which was reduced to 475,081 autosomal and 815
303 mitochondrial variants after quality filtering. The final dataset comprises 153 samples with mean read
304 depths of 520.7x (range: 251.4-998.2x) and 66.0x (range: 14.8-726.2) at mitochondrial and autosomal
305 loci. Location and sequence coverage statistics for all samples in the final dataset are listed in
306 Supplemental Table 1.
307
308 *Summary statistics* – Autosomal and mitochondrial summary statistics for π, *H*, Tajima's *D*, Θ, and *Ne*
309 are shown in Table 1. π, Θ, and *Ne* are similar between the West African and Brazilian *S. mansoni*
310 populations but are 2-3 times lower than observed in Tanzanian *S. mansoni*. Tajima's *D* values range
311 between slightly positive to negative (Tajima's *D* = -1.417 – 0.034) in *S. mansoni* (Table 1; Supplemental
312 Table 2). All the African populations show negative Tajima's *D*, consistent with natural selection or
313 population expansion. However, the Brazilian Tajima's *D* is positive, which is inconsistent with a
314 bottleneck during colonization of South America. Mitochondrial diversity was quantified with π and
315 haplotype diversity (*H*). *H* in all populations is very high (<0.978) indicating that all mitochondrial
316 haplotypes are unique. Mitochondrial π follows the same pattern as autosomal π with the exception that
317 the mitochondrial π in Brazil is lower than may be expected when compared to measures in Niger and
318 Senegal. $F_{ST}$ values between *S. mansoni* populations are shown in Table 2 and were highest in pairwise
319 comparisons that included the Tanzanian population. Mantel tests showed significant signs of isolation-
320 by-distance within Africa ($r$ = 0.64, $p$ = 0.001) and in African and Brazilian ($r$ = 0.77, $p$ = 0.001) *S.*
321 *mansoni* samples.
322
323 Diversity and *Ne* values were lower in *S. rodhaini* than all *S. mansoni* populations but should be
324 interpreted with caution. Eight of the nine *S. rodhaini* samples came the same laboratory-maintained
325 population. $F_{ST}$ between *S. rodhaini* and individual *S. mansoni* populations ranged from 0.844 (*S.*
326 *rodhaini* vs. Sm Tanzania) - 0.937 (*S. rodhaini* vs. Sm Senegal).

Figure 3 shows linkage disequilibrium (LD) decay (binned and smoothed) from pairwise $r^2$ values. LD decays to $r^2 \leq 0.2$ within 500Kb for all populations. LD was weakest in the Tanzania population with $r^2$ decaying to 0.5 in 28 bp. LD decay in the other three populations (Senegal, Niger, and Brazil) was relatively consistent with LD decaying by half ($r^2 = 0.5$) between 15,150 bp (Niger) and 26,196 bp (Brazil). Both Senegal and Niger show high levels of LD even between SNVs on different chromosomes: this results from lower sample size (Senegal, n=25; Niger, n=10) in these two populations.

*Admixture with S. rodhaini.* We asked whether hybridization with *S. rodhaini*, a closely-related schistosome infecting rodents, might contribute to the high genetic diversity observed in East Africa vs West Africa/south American *S. mansoni*. To investigate this, we used three statistics (D, $D_3$, and $F_3$) to test for admixture between *S. mansoni* and *S. rodhaini*, with particular emphasis on the Tanzanian populations of *S. mansoni*. Each of these statistics, attempts to identify the presence of admixture in different ways. D and $D_3$ values $\neq 0$ indicate admixture and positive and negative values determining the direction of introgression. $F_3$ values $< 0$ indicate admixture between the two source populations. None of the three statistics, or any of the population combinations, returned values containing significant signals for admixture (Table 3). These results suggest that hybridization/introgression between *S. mansoni* and *S. rodhaini* may make no detectable contribution to elevated diversity in East Africa.

*Population structure* - We examined population structure using PCA and ADMIXTURE with 38,197 unlinked autosomal SNVs. Two PCAs were generated, with and without the *S. rodhaini* outgroup (Figure 4). The two species were differentiated along PC1 (34.7% variance) when *S. rodhaini* was included (Figure 4A). *S. mansoni* samples cluster into geographically defined groups when *S. rodhaini* is excluded (Figure 4B). East African samples were distinct from all other *S. mansoni* samples, except for one whole genome sample collected in Kenya (see below). Samples from the Americas, including those from the Caribbean and Brazil, showed a closer relationship with Cameroon and Nigerien samples than those from Senegal.

We used ADMIXTURE to assign individuals to one of *k* populations, where *k* is between 1 and 20 (Figure 5). Cross-validation scores (Evanno et al., 2005) were minimized when *k* was 4 or 5. Both *k* = 4 or 5 split *S. mansoni* samples into geographically defined populations with two major differences. First, *k*=4 showed that the allelic component primarily associated with Brazil was found at moderate levels in Cameroon and Nigerien individuals. Second *k*=5 split the West African samples into a Senegalese and a Cameroonian + Nigerien populations.

As observed in the PCA, the Kenyan , whole-genome sample contained a large portion of alleles associated with samples from the Americas (~40-60%) in the ADMIXTURE analysis. Crellen *et al*. (2016) recovered similar results and hypothesized that the Kenyan sample may be reflecting human-trafficking routes between Portuguese and Arab slave traders out of the port of Mombasa. Given that only a single sample is available from this region, and contamination with South American strains during laboratory passage is a possible alternative explanation, we chose to remove the Kenyan sample from downstream analyses.

*Phylogenetics* – We used three different phylogenetic methods to investigate the evolutionary relationships between sequences (Figure 6). First, we generated median-joining network from 815 mitochondrial SNVs of which 477 where phylogenetically informative (Figure 6A). The haplotype network identified three major haplotypes roughly corresponding with the geographic partitioning of the samples. The haplogroups include an East African clade (Tanzania, Uganda), a Senegal group, and intermediate haplogroup with samples from the Brazilian and Nigerien populations. Caribbean samples were not assigned to a single haplogroup. The single sample from Puerto Rico was associated with the major Brazilian and Nigerien haplotype, and the two Guadalupe samples tended to be more strongly associated with Senegalese haplotypes. Additionally, the sample from Cameroon was only single step removed from the most common Brazilian haplotype.

A coalescent-based species tree from 100,819 parsimony informative SNVs was generated with SVD-quartets (Figure 6B). Quartet sampling was limited to 100k quartets which sampled 0.43% of all distinct quartets present in the alignment. The final species tree was consistent with 84.7% of all the quartets sampled. Unlike the mitochondrial tree, samples fall into well supported clades corresponding with geography with two exceptions. Samples from East Africa and Niger formed independent paraphyletic clades. In both cases paraphyly was induced by a single individual. Bootstrap support was generally higher in the quartet species tree than in the mitochondrial tree. West African samples formed a well-supported monophyletic clade with the Brazilian and Caribbean samples, indicating a shared origin for these parasites. Brazilian and Caribbean samples appear to have a polyphyletic origin within a larger clade containing West African parasites.

Finally, migration between populations was quantified with TreeMix (Figure 6C). We only examined populations with more than 5 individuals which excluded Cameroonian and Caribbean samples from the analysis. The topologies linking the remaining populations (East Africa, Niger, Senegal, and Brazil) in the TreeMix and species tree were identical. The species tree was slightly improved with the

addition of a single migration edge from Brazil to Niger. This migration edge significantly improved the likelihood score of the topologies from 73.6026 to 73.7165 with an edge weight of 0.1 ($p = 0.0081$).

*Selection* – We used msprime to generate a set of neutrally evolving SNVs based on parameters specific to each of the sampled populations. These neutrally evolving SNVs were distributed across an 88.9 Mb chromosome that was equal in size to *S. mansoni* chromosome 1 (HE601624.2). We then transposed the HE601624.2 exome annotation onto the simulated chromosome to extract "exome" data. This process was repeated 342 times per population to produce a set of neutrally evolving loci to use as controls when examining selection on actual samples. We used these neutral simulations to generate maximum (100%) threshold values for H-Scan and Sweepfinder2 test statistics expected under neutrality (see below). The mean ($\bar{x}$) simulated SNV count across all replicates was: $\bar{x}_{Brazil}$ = 11,198 (range = 10,874-11,552), $\bar{x}_{Niger}$ = 8,448 (range = 8,141-8,706), $\bar{x}_{Senegal}$ = 11,977 (range = 11,642-12,309), $\bar{x}_{Tanzania}$ = 33,168 (range = 32,575-33,741). The number of observed SNVs was between 72.2-89.7% of the mean number of simulated SNVs (Brazil = 8,947; Niger = 6,107; Senegal = 9,381; Tanzania = 29,765). The reduced number of SNVs in the neutral data is likely due to the absence of selection that is presumed to be acting on the exome data sets.

The H-Scan and SweepFinder2 results are shown in Figure 7. *pcadapt* results are show in Supplemental Figure 1. Each program uses different methodology to detect sites under selection. H-Scan calculates *H* with homozygous tract length and the number of haplotypes to identify genome regions that have undergone selective sweeps. *H* values were highly variable for each population, even within windows smaller than 100Kb. In Tanzania only 2 of the 475,081 SNVs had *H* values higher than were generated from neutral simulations. SweepFinder2 calculates deviations from a neutral site frequency spectrum correcting for the possibility of background selection. A likelihood ratio (LR) describes the probability of positive selection vs. neutral evolution and background selection within a designated window. Sweepfinder2 was able to clearly define multiple peaks for each population when compared to H-Scan. Each of the four populations had regions greater than neutral expectations. *pcadapt* identifies SNVs significantly associated with population differentiation. In this analysis we found 442 SNV outliers, after multiple test correction, that were distributed across all seven autosomes at 127 loci. These regions were distributed across 280.2 Mb. $F_{ST}$ of the *pcadapt* outliers ($F_{ST}$ = 0.544) was significantly higher than in the remaining population outliers ($F_{ST}$ = 0.195).

We defined "putative regions of selection", as those that have most likely experienced positive selection. These regions contain variants (i) with both *H* and LR values in the 99$^{th}$ percentile (ii) are

greater than the neutral thresholds and (iii) have a signal of population-specific directional. All SNVs meeting one or more of these criteria are listed in Supplemental Table 4.

Our results recovered 5, 3, and 3 putative selected regions in Brazil, Niger, and Senegal respectively (Figure 7; Supplemental Table 5). Information regarding the number of regions, SNVs, and genes identified are presented in Table 4 and Table 5. $\pi$ (Supplemental Figure 2) and Tajima's *D* (Supplemental Figure 3) were depressed in these regions compared to genome-wide values (Supplemental Table 3) which is consistent with loci experiencing selection. On average the size of each region was relatively small (1,395,643 bp) and in two instances, these sites were shared between populations. The Brazilian and Senegalese populations shared a site on chromosome 3 at HE601626.2:30,092,830:31,936,551, while the Nigerien and Senegalese populations shared a peak on chromosome 4 at HE601627.2:31,216,154:32,138,352. We did not recover any regions or sites of interest in the Tanzanian population in large part because only two of 475,081 sites had higher *H* values the largest *H* from neutral simulations (*H* = 45532.9). These variants are on chromosome 3 at HE601626.2:6,500,422 and HE601626.2:6,503,809 and are adjacent to each other in our filtered SNV dataset.

We identified 116-157 genes within "putative selected regions" in the Brazilian, Nigerien, and Senegalese populations (Supplemental Table 6). Within these populations 10, 5, and 7 genes contain SNVs meeting the 99$^{th}$ percentile. Several genes identified in these regions were shared between populations. Brazil and Senegal shared 48 genes in target regions, and Senegal and Niger shared 22 genes in target regions. Three genes with 99$^{th}$ percentile SNVs were shared between populations: Smp_313490 and Smp_167890 (Niger and Senegal), Smp_123520 (Brazil and Senegal; Table 5).

## Discussion

We examined the impact of human-mediated dispersal of *S. mansoni* during the Trans-Atlantic slave trade. Previous analyses with *S. mansoni* used mitochondrial data or had limited sampling from the Americas (Crellen et al., 2016; Morgan et al., 2005; Webster et al., 2013). To further these studies, we included 135 exome sequences available from natural populations in Brazil (n=45), Niger (n=10), Senegal (n=25), and Tanzania (n=55) which we combined with the existing genome sequences from Cameroon (n=1), the Caribbean (n=4), Senegal (n=1), and Uganda (n=2) (Berriman et al., 2009; Chevalier et al., 2016; Crellen et al., 2016). We used these data to examine *S. mansoni* expansion across Africa, explore examine parasite colonization of the Americas, quantify signatures of selection during colonization, and detect hybridization with *S. rodhaini*, a closely related parasite utilizing a rodent host.

*Elevated East African diversity and* S. mansoni *expansion across Africa* – A striking result from this study is the dramatic reduction in genetic diversity between East and West Africa. Sequence summary statistics indicate that the East African population has 2-3-fold greater nucleotide diversity ($\pi$), larger *Ne* and greater mitochondrial diversity than the other populations (Table 1; Figure 6A). Phylogenetic analyses rooted with *S. rodhaini* clearly indicate that the East African *S. mansoni* samples from Tanzania and Uganda are sister to a clade containing all other *S. mansoni*, including those from the Caribbean, Brazil, Senegal, Niger, and Cameroon. Our data supports previous work from whole genomes, mitochondrial genes, and microsatellites suggesting that *S. mansoni* emerged in East Africa (Crellen et al., 2016; Morgan et al., 2005; Webster et al., 2013). In addition, our estimates of *Ne* are in within a range book-ended by other whole genome studies (Berger et al., 2021; Crellen et al., 2016).

The rapid decay in LD observed in E. Africa compared with W. African and American populations provides further evidence that East African *S. mansoni* populations are ancestral. Similar reductions in rate of LD decay have been observed in humans and malaria parasites, outside of their ancestral Africa range (Anderson et al., 2000; Gurdasani et al., 2015; Neafsey et al., 2008). Rapid breakdown in LD also has important practical applications for genome wide association analyses, because it allows mapping of phenotypic traits to very narrow regions of the genome (Mackay & Huang, 2018). There are multiple biomedically important traits of interest that vary in *S. mansoni* populations, including drug susceptibility or resistance, host-specificity and cercarial production (Anderson, LoVerde, Le Clec'h, & Chevalier, 2018). We recently used GWAS for mapping resistance to the first line drug (Praziquantel) in laboratory schistosome populations (Le Clec'h et al., 2021). Most laboratory *S. mansoni* populations tested were from S. America, where LD decays relatively slowly: these are not ideal for GWAS. Establishment of laboratory *S. mansoni* populations from East Africa, or GWAS analyses using parasites directly from the field would be valuable for future GWAS with *S. mansoni*.

There is minimal allele sharing or migration between East African and other *S. mansoni* populations. $F_{ST}$ comparisons that include Tanzania are greater than other comparisons, (with Tanzania - $F_{ST} = 0.355$; excluding Tanzania - $F_{ST} = 0.206$). East African populations are among the most strongly differentiated populations in the PCA analyses (Figure 4B) and the East African population component in ADMIXTURE analyses is absent, or at minimal levels, in other *S. mansoni* populations. Finally, mitochondrial haplotypes in Uganda and Tanzania form a distinct haplogroup from other *S. mansoni* populations. These data indicate that, while E. Africa is the likely origin of *S. mansoni*, migration or allele sharing between E. Africa and other populations is restricted.

We do not expect that human movement is a major barrier between East and West African schistosome populations. However, differences in snail – schistosome compatibility in East and West

Africa may provide barriers to gene flow. Sympatric host–parasite combinations tend to show greater compatibility than allopatric combinations. This is seen in multiple host parasite systems including *Daphnia*–microsporidia (Ebert, 1994), trematode infections of snails (Lively, 1989) and minnows (Ballabeni & Ward, 1993). Strong host-specificity exists within the *Biomphalaria* and *S. mansoni* system (Mitta et al., 2017; Theron et al., 2014; Webster & Woolhouse, 1998) and a review shows that compatibility is greater between sympatric *Biomphalaria-S. mansoni* combinations (Morand, Manning, & Woolhouse, 1996). Sympatric schistosome–snail combinations result in rapid immune suppression and rapid parasite development, while allopatric schistosome-snail combinations result in a slower immune cell proliferation and a non-specific generalized immune response which reduced parasite growth and establishment (Portet et al., 2019). There is a developing understanding of *Biomphalaria* phylogenetics (Jorgenson, Kristensen, & Stothard, 2007), phylogeography (Dejong et al., 2003), and compatibility relationships among East African snail species (*B sudanica*, *B. pffeiferi* and *B. choanomphala*) and *S. mansoni* (Mutuku et al., 2021; Mutuku et al., 2017), but further research is needed to understand compatibility of allopatric snail–schistosome combinations from East and West Africa. We suggest that the presence of fine-scale geographic structure of *Biomphalaria* populations (Webster et al., 2001) and local adaptation in sympatric *Biompahalaria*-schistosome combinations may limit parasite geneflow between E. African and W. Africa.

PCA analyses differentiate *S. mansoni* populations on an East-to-West gradient along PC2 (Figure 4B). Mitochondrial haplotypes in Senegal and Cameroon are intermediate to Tanzania and Senegal (Figure 6A). The species tree from autosomal SNV data (Figure 6A) indicates that *S. mansoni* is in a series of nested, well-supported clades from East Africa (Tanzania + Uganda) to Cameroon to West Africa (Niger + Senegal). $F_{ST}$ and *r* (Mantel) values between Tanzania, Niger, and Senegal reflect increasing isolation with distance across Africa. These observations combined with the origination of *S. mansoni* in East Africa confirms an East-to-West, stepwise expansion of *S. mansoni* from Tanzania and Uganda → Cameroon → Niger → Senegal (Crellen et al., 2016; Morgan et al., 2005; Webster et al., 2013).

*Does Hybridization between* S. rodhaini *and* S. mansoni *contribute to elevated East African Diversity* – Several closely related *Schistosoma* species are able to hybridize with the production of viable offspring confirmed via experimental rodent infections. The potential for hybridization between animal and human *Schistosoma* species is a significant public health concern (Borlase et al., 2021; Léger et al., 2020; Leger & Webster, 2017; Stothard, Kayuni, Al-Harbi, Musaya, & Webster, 2020). Our group, and others, have recently shown that ancient hybridization and adaptive introgression has resulted in the

transfer of genes from the livestock species *Schistosoma bovis* into *S. haematobium*: west African *S. haematobium* genomes contain 3-8% introgressed *S. bovis* sequences and *S. bovis* alleles have reached high frequency in some genome regions (Platt et al., 2019; Rey, Toulza, et al., 2021). The sister species of *S. mansoni*, *S. rodhaini*, parasitizes rodents and is primarily located in Eastern Africa (Rey, Webster, et al., 2021). *S mansoni* and *S. rodhaini* have been shown to readily hybridize in and produce fertile offspring in the lab (Théron, 1989). Natural hybrids have been reported in Kenya and Tanzania (Morgan et al., 2003; M. Steinauer et al., 2008), although hybrids have only been detected from their snail intermediate host and never encountered in the mammalian hosts humans and rodents (Rey, Webster, et al., 2021). We were unable to find evidence of hybridization in 55 samples collected from Tanzania. Both species are clearly separated in genotypic space with differences between the species accounting for the largest component in the PCA (Figure 4A; PC1=34.7% variation) and $F_{ST}$ between *S. rodhaini* and *S. mansoni* populations is very high ($F_{ST}$ = 0.912). Admixture analyses also clearly differentiated *S. rodhaini* from all other *S. mansoni* populations (Figure 5) and we were unable to identify admixture signal between *S. rodhaini* and the Tanzanian population with genome-wide statistics including D, $D_3$, or $F_3$, (Table 3).

Hybridization between these two species is thought to be rare (≤7.2%) (Morgan et al., 2003; Rey, Webster, et al., 2021; M. Steinauer et al., 2008; M Steinauer et al., 2008). Our sample size may not be large enough to identify rare hybrids. Further we analyzed exome (coding) data, which may underrepresent introgressed alleles if they are selected against. These caveats aside, our analyses clearly fail to identify recent hybridization between *S. mansoni* and *S. rodhaini*. We conclude that *S. rodhaini* introgression does not contribute to the high, genetic diversity in our Tanzanian *S. mansoni* samples.

*Expansion into the Americas* – Previous work has shown that *S. mansoni* was exported from Africa to the Americas during the Trans-Atlantic slave trade (Crellen et al., 2016; Després et al., 1993; Files, 1951; Fletcher et al., 1981; Morgan et al., 2005; Webster et al., 2013). Here, we use genomic data to investigate the likely source population(s), number of introductions, evidence for bottlenecks and parasite adaptation during colonization.

Source populations. Of the two West African populations sampled (Niger and Senegal), our results support stronger relationships between Brazil and Niger, than with Senegal. While the species tree (Figure 6B) appears to rule out Niger or Senegal as the direct source population for Brazilian *S. mansoni*, there is evidence of allele sharing between the Nigerian and Brazil populations. First, the dominant mtDNA haplotype in Brazilian and Nigerien samples are shared (Figure 6A). Second, every Nigerien

554 population contains at least 10.1% of the Brazilian component as shown in the Admixture analyses (mean
555 15.8%; Figure 5). Third, Niger and Brazil are more closely associated with each other in the genotypic
556 continuum represented by PC2 than Brazil is to other African populations (Figure 4B). Finally, TreeMix
557 identified a single weak migration edge between Brazil and Niger (Figure 6C) confirming a relationship
558 between these two populations.

559 A simple hypothesis from the data is that, assuming the general East-to-West expansion holds at
560 finer geographic scales, the source population that was eventually exported to Brazil, likely occurs
561 somewhere between Benin and Angola. These countries fell within the Bight of Benin, Bight of Biafra,
562 and West Central Africa slave trading regions (Figure 8). Our Brazilian samples were collected in Ponto
563 dos Volantes in Minas Gerais, Brazil. This location is relatively equidistant from major slave ports in
564 Bahia (527 Km) and around Rio de Janeiro (706 Km). In all, more than 3.5 million (M) enslaved peoples
565 were imported into Brazil (Supplemental Table 7; Slave Voyages Database, 2009). Of all the enslaved
566 people exported to Brazil, 82% disembarked at ports in either Bahia (1.3M) or Rio de Janerio (1.5M). In
567 Africa, slave exporting markets in the Bights of Benin and Biafra and West Central Africa were
568 responsible for 53.6%, 5.0% and 33.8% of peoples exported to Bahia and 1.3%, 1.1% and 74.9% of
569 peoples exported to south east Brazilian ports (Slave Voyages Database, 2009). Taken together these
570 data imply that the Brazilian population we sampled in Ponto dos Volantes most likely originated from
571 markets in the Bight of Benin or West Central Africa. In our phylogenetic analyses, the Brazilian
572 population falls within a clade containing samples from Senegal, Niger, and the Caribbean but excludes
573 the single Cameroonian sample. If the Cameroonian sample is representative of the Cameroonian
574 population, then it may be that Cameroon and Niger represent the eastern and western limits of the
575 unrepresented source population. This area more closely aligns with the Bight of Benin, a region
576 containing parts of Nigeria and Benin. Additional samples from these regions are needed to test this
577 hypothesis. It is also important to note that, the Brazilian samples here represent a single, geographic
578 location (Ponto dos Volantes) and that the source for this population may not extrapolate to larger regions,
579 or even to other locations in Brazil.
580
581 No evidence for population bottlenecks during colonization. Previous mtDNA analyses have
582 shown reduced diversity in S. American parasites (Després et al., 1993; Fletcher et al., 1981; Morgan et
583 al., 2005; Webster et al., 2013) suggesting bottlenecks during colonization. Our data is consistent with
584 this showing a 2-3-fold reduction in mtDNA in Brazil compared to W. African parasites. However,
585 genome-wide summary statistics of autosomal sequence data tell a different story, and provide no
586 evidence for population bottlenecks associated with *S. mansoni* introduction and establishment in Ponto

dos Volantes, Brazil (Table 1). Nucleotide diversity, measured by π, was higher in Brazil than either of the West African populations (Niger or Senegal), perhaps because the Brazilian population is derived from several west African source populations. The Brazilian $N_e$ was roughly the same as Niger and comparable to Senegal. Mean Tajima's *D* calculated from genome-wide, SNV data is negative in each of the African populations and highest in the Caribbean samples (mean Tajima's $D = 0.929$; Supplemental Table 2). Tajima's *D* values in our study using exome SNV data from different regions of Niger, Senegal and Tanzania were comparable with values calculated with whole genome SNV data from Ugandan *S. mansoni* (Berger et al., 2021). By contrast mean Tajima's *D* in Brazilian samples is close to 0 (mean Tajima's $D = 0.034$; 95% CI [-0.18 to -.085]) and is significantly greater than mean Tajima's *D* in the African samples (T-Test $p<0.001$; Supplemental Table 2).

The nuclear genomic data clearly suggest that the establishment of *S. mansoni* in Brazil was not associated with a significant population bottleneck and had minimal impacts on genome-wide levels of genetic diversity. The discrepancy between mtDNA and nuclear DNA may stem from two sources. First, mtDNA has an effective population size ¼ that of nuclear genes (Birky, Maruyama, & Fuerst, 1983), and can potentially provide a more sensitive indicator of bottlenecks. Second, and perhaps more critical, mtDNA constitutes a single marker, so may poorly reflect population history (Anderson, 2001). Extensive laboratory passage may also result in bias in population summary statistics. For example, the Caribbean samples examined have undergone 2-15 generations of lab passage which is probably responsible for the elevated Tajima's *D* at in this population (Crellen et al., 2016).

Number of introductions. All Brazilian and Caribbean samples are paraphyletic fall between the Cameroonian sample and West African clade in Figure 6B. The relationships among these samples are resolved but not supported outside of a monophyletic clade containing all Brazilian samples. As a result, the autosomal phylogeny by itself does not conclusively support one or multiple introductions into the Americas. The two Caribbean samples from Guadeloupe island contain unique mitochondrial haplotypes absent from Brazilian (Figure 6A); however, this is, at best, only weak evidence for independent introductions into Brazil and the Caribbean from the data we have available. The higher autosomal diversity of Brazilian *S. mansoni* compared with two sampled West African populations provides additional indirect evidence for multiple origins.

*Plasmodium falciparum* and *Wuchereria bancrofti* became established in the Americas at the same time as *S. mansoni*, and without apparent bottlenecks (Small et al., 2019; Yalcindag et al., 2012). In both cases it is hypothesized that high levels of diversity were maintained by recurring introductions from multiple sources (Rodrigues et al., 2018; Small et al., 2019; Yalcindag et al., 2012). More than 3.5

M enslaved peoples were exported into Brazil from across the continent of Africa (Supplemental Table 7; Slave Voyages Database, 2009). *S. mansoni* prevalence varies widely between sites in western and central Africa, with estimates ranging from 0.3% (Gambia; Sanneh et al., 2017) to 89% (Democratic Republic of the Congo; Kabongo et al., 2018). Assuming *S. mansoni* prevalence during the period of the Atlantic Slave Trade is comparable to current levels and that infected persons contained multiple *S. mansoni* individuals (genotypes; Van den Broeck et al., 2014), it is plausible that hundreds of thousands, and more likely millions, of reproductively viable *S. mansoni* were introduced into Brazil from across Africa. As a result, individual genotypes that may have been separated by thousands of miles in Africa were brought into close contact in the Americas and would lead to higher levels of diversity in there than in any single population in Africa.

Adaptation during colonization – We hypothesized that *S. mansoni* introduced into the Americas would have been exposed to novel selective pressures as they adapted to new biotic and abiotic challenges. For example, the *S. mansoni* life cycle requires an intermediate snail host in which miracidia mature into cercariae that are capable of infecting humans. In Africa, *S. mansoni* use *Biomphalaria alexandrina*, *camerunensis*, *choanomphala*, *pfeifferi*, *stanleyi*, and / or *sudanica* as the snail host, none of which are present in the Americas. Instead, these parasites have adapted to using different *Biomphalaria* hosts, including *B. glabrata*, *B. straminea*, and *B. tenagophila* (Figure 2; Hailegebriel, Nibret, & Munshea, 2020; Kengne-Fokam, Nana-Djeunga, Bagayan, & Njiokou, 2018; Vidigal et al., 2000). We examined exomic SNV data to identify genes and larger regions of the genome under selection at a finer scale and identified 0-5 putative regions of selection from each of the major populations (Table 4).

In the Brazilian samples, we identified five putative selected regions that contain 126 genes (Supplemental Table 6). $\pi$ and Tajima's *D* was significantly reduced in these five regions compared to genome wide averages which is expected if these loci are, or have been, under selection (Supplemental Table 3). One region is shared between the Brazilian and Senegalese populations. Forty-six genes fall within this Sengal:Brazil overlapping region, leaving 80 genes and 4 loci that are likely experiencing population specific, positive selection specific. Even within the group of 80 genes, there are 9 with strong signals of selection (Table 5). These genes contained variants with *H* and LR values in the 99$^{th}$ percentile in addition to being greater than the threshold defined by neutral simulations. Several genes within this group are associated with housekeeping functions, including transcription and protein degradation (Smp_060090, Smp_162000, and Smp_246630). Two uncharacterized proteins were identified (Smp_341570 and Smp_123590) but we are not able to speculate on their function. Of particular interest

653 are two possible transcription factors, an uncharacterized protein containing a *helix, loop, helix* domain
654 (Smp_123570) and a putative TATA-box binding protein (Smp_073680). It is possible that adaptation to
655 the Brazilian environment was driven by changes in gene expression, however more work is needed to
656 understand the potential role of these loci in adaptation to the Americas.
657
658 *Selection on African S. mansoni* – We examined selection on African *S. mansoni* as part of the
659 process to identify unique signals of selection in the Brazilian population. We identified 112 and 157
660 genes under selection in 3 regions each for the Nigerien and Senegalese populations (Table 4). One of
661 the three regions, and 22 genes, were shared between Niger and Senegal. We failed to identify any regions
662 of selection in Tanzania using our combined criteria, however results from individual tests of selection
663 (H-Scan, SweepFinder2) did overlap at nine of 25 regions identified in large Ugandan by Berger et al.
664 (2021) (Supplemental Table 8). These regions were identified using a variety of within (iHS) and
665 between population (FST, XP-EHH) tests on miracidia isolated from two Kenyan locations with differing
666 histories of praziquantel treatment.
667
668 **Conclusions**
669 Our analyses identified an East-to-West expansion of *S. mansoni* across Africa. Sometime during
670 this expansion one or more Central African population(s), likely located between Angola and Benin, were
671 transported to Brazil. Genome-wide signatures of diversity, measures of allele frequencies (Tajima's *D*),
672 and estimates of *Ne* are comparable between *S. mansoni* in Brazil, Niger, and Senegal and do not imply
673 the presence of a bottleneck during the establishment of *S. mansoni* in Brazil. We did find five genome
674 regions under selection in Brazil, 4 of which are population specific. In total, 80 genes fall within these
675 regions and may be associated with *S. mansoni*'s adaptation to novel selection pressures associated with
676 the Americas. We identified 9 genes with the strongest signals of selection that are candidates for future
677 experimental work.
678
679 **Acknowledgments**
680 Sandra Smith and John M. Heaner (Texas Biomedical Research Institute) provided computational
681 research support. Matthew Berriman provided the most recent *S. mansoni* genome sequence and
682 annotation prior to publication. This research was funded by the National Institute of Allergy and
683 Infectious Diseases (NIAD R01 AI097576-01 and NIAD 5R21AI096277-01), Texas Biomedical
684 Research Institute Forum (award 0467). Samples acquisition was supported by funding from the
685 Wellcome Trust (104958/Z/14/Z), the Gates Foundation coordinated by the University of Georgia

Research Foundation Inc. (RR374-053/5054146 and RR374-053/4785426), and a ZELS research grant (combined BBSRC, MRC, ESRC, NERC, DSTL, and DFID: BB/L018985/1). The authors declare no competing interests.

## Author Contributions

Study design done by F.D.C., T.J.A., and W.L.C; formal analysis done by F.D.C, R.N.P., and T.J.A.; writing, original draft, done by R.N.P. and T.J.A.; reviewing and editing, done by A.G., A.E., B.W., D.R., F.D.C., G.O., J.P.W., M.M, P.T.L., R.R.dA., R.N.P., S.K., T.J.A., and W.L.C.; investigation was done by. F.D.C., M.M., and W.L.C.; resources were provided by: F.D.C., and T.J.A.

## Data Availability

Data used in this manuscript was previously published in (Berriman et al., 2009; Chevalier et al., 2019; Crellen et al., 2016; International Helminth Genomes Consortium, 2019; Le Clec'h et al., 2021) under multiple NCBI BioProject (PRJNA439266, PRJNA560070, PRJEB522, PRJEB526, PRJNA743359 and PRJNA773498) and NCBI Short Read Archive (ERR046038, ERR103049, ERR103050, ERR119614, ERR119615, ERX284221, ERR310938, ERR539846, ERR539847, ERR539848, and ERR9974) accessions.

## Code Availability

Scripts, notebooks, and environmental yaml files are available at https://github.com/nealplatt/sch_man_nwinvasion/releases/tag/v0.2 (last accessed 21 Oct 2021) or https://doi.org/10.5281/zenodo.5590460 (last accessed 21 Oct 2021).

Table 1. Whole genome summary statisitics for *Schistosoma mansoni* populations and *S. rodhaini*

| | **n** | ***π*** | ***π* (mito)** | ***H* (mito)** | **Tajima's *D*** | **Θ** | ***Ne*** | ***π* (PRS)** | **Tajima's *D* (PRS)** |
|---|---|---|---|---|---|---|---|---|---|
| *S. rodhaini* | 9[a] | 5.61E-04 | Na[b] | Na[b] | 0.479 | 4.29E-04 | 13,226 | Na[c] | Na[c] |
| Sm (Brazil) | 45 | 6.93E-04 | 2.10E-03 | 0.984 | 0.034 | 5.84E-04 | 18,032 | 2.22E-04 | -1.218 |
| Sm (Niger) | 10 | 6.00E-04 | 5.89E-03 | 0.978 | -0.579 | 6.07E-04 | 18,737 | 2.46E-04 | -1.018 |
| Sm (Senegal) | 25 | 4.97E-04 | 4.72E-03 | 0.997 | -1.417 | -1.375 | 21,992 | 1.75E-04 | -1.816 |
| Sm (Tanzania) | 55 | 1.45E-03 | 7.25E-03 | 1.0 | -0.729 | -0.739 | 51,508 | Na[c] | Na[c] |

"Sm" - *Schistosoma mansoni*, "n" - number of samples, "π" - nucleotide diversity, "*H*" - haplotype diversity, "Θ" - Watterson estimator, "*Ne*" - effective population size, "PRS" - putative region of selection

a) 8 of 9 *S. rodhaini* samples came from a single lab population: population statistics are likely biased

b) No *S. rodhaini* reads mapped to the *S. mansoni* mitochondria.

c) Not calculated.

Table 2. $F_{ST}$ between Schistosoma species and populations

| **Pop1** | **Pop2** | **$F_{ST}$** | **S.E.** |
|---|---|---|---|
| *S. rodhaini* | Sm (Caribbean) | 0.929 | 0.0012 |
| *S. rodhaini* | Sm (Tanzania) | 0.844 | 0.0016 |
| *S. rodhaini* | Sm (Senegal) | 0.937 | 0.0013 |
| *S. rodhaini* | Sm (Niger) | 0.931 | 0.0011 |
| *S. rodhaini* | Sm (Brazil) | 0.919 | 0.0013 |
| Sm (Caribbean) | Sm (Tanzania) | 0.279 | 0.0036 |
| Sm (Caribbean) | Sm (Senegal) | 0.323 | 0.0085 |
| Sm (Caribbean) | Sm (Niger) | 0.236 | 0.0071 |
| Sm (Caribbean) | Sm (Brazil) | 0.154 | 0.0067 |
| Sm (Tanzania) | Sm (Senegal) | 0.416 | 0.0032 |
| Sm (Tanzania) | Sm (Niger) | 0.348 | 0.0031 |
| Sm (Tanzania) | Sm (Brazil) | 0.379 | 0.0034 |
| Sm (Senegal) | Sm (Niger) | 0.135 | 0.0042 |
| Sm (Senegal) | Sm (Brazil) | 0.235 | 0.0047 |
| Sm (Niger) | Sm (Brazil) | 0.152 | 0.0036 |

Abbreviations: "Sm" - *Schistosoma mansoni*; "S.E." - standard error

Table 3. Admixture statistics

| **Comparison** | **Test Stat** | **SE** | ***Z*** |
|---|---|---|---|
| D (Patterson *et al*. 2012) | | | |
| (((Br,Tz)Sr),Smr) | -0.033 | 0.0404 | -0.807 |
| (((Ni,Tz)Sr),Smr) | 0.006 | 0.0437 | 0.129 |
| (((Se,Tz)Sr),Smr) | 0.051 | 0.0473 | 1.070 |
| $F_3$ (Patterson *et al*. 2012) | | | |
| Br: Tz, Sr | 0.764 | 0.0207 | 36.910 |
| Ni: Tz, Sr | 0.993 | 0.0299 | 33.214 |
| Se: Tz, Sr | 1.265 | 0.0461 | 27.428 |
| $D_3$ (Hahn and Hibbens 2019) | | | |
| (Br, Tz, Sr) | -0.0047 | 0.00036 | 0.418 |
| (Ni, Tz, Sr) | -0.0046 | 0.00034 | 0.419 |
| (Se, Tz, Sr) | -0.0054 | 0.00048 | 0.353 |
| (Cm, Tz, Sr) | 0.0005 | 0.00066 | -0.022 |
| (Cr, Tz, Sr) | -0.0048 | 0.00029 | 0.517 |
| (Ug, Tz, Sr) | 0.0034 | 0.00062 | -0.174 |

SE - standard error, "*Z*" - *Z*-score of the mean from 0. Population abbreviations: "Br" - Brazil, "Cm" - Cameroon, "Cr" - Caribbean, "Ni" - Niger, "Se" - Senegal, "Smr" - *Schistosoma margrebowiei*; "Sr" - *Schistosoma rodhaini*; "Tz" - Tanzania; "Ug" - Uganda

711

Table 4. Number of SNPs and regions indentified in genome-wide scans for selection

| | **Brazil** | **Niger** | **Senegal** | **Tanzania** |
|---|---|---|---|---|
| Outlier SNVs (H-Scan) | 4,366 | 44,024 | 9,182 | 2 |
| Outlier SNVs (Sweepfinder) | 12,871 | 41,525 | 23,379 | 4,288 |
| Merged Regions (1/3 Mb) | 113 | 250 | 168 | 46 |
| 99th Percentile SNPs | 703 | 239 | 176 | 0 |
| Outlier SNPSs (pcadapt) | | 442 | | Na |
| Putative regions under selection | 5 | 3 | 3 | 0 |
| Genes in selected regions | 116 | 112 | 157 | 0 |
| Genes w/99th Perc SNPs | 10 | 5 | 7 | 0 |

*Outlier SNPS are those that are greater than the maximum value derived from neutrally simulated data

712

Table 5. Genes in *Schistosoma mansoni* populations with the strongest signals of directional selection

| Gene ID | UniProtKB Accession | UniProtKB Description | HHsearch Annotation |
|---|---|---|---|
| **Brazil** | | | |
| Smp_060090 | G4VAJ9 | 40S ribosomal protein S12 | Ribosomal protein S12e Ribosome |
| Smp_073680 | G4V701 | Putative TATA-box binding protein | DNA-directed RNA polymerase II subunit |
| Smp_123510 | A0A5K4EKW3 | Vacuolar protein sorting-associated protein 16 homolog | Vps16_C |
| Smp_123520[1] | A0A3Q0KKK4 | Putative RNA $M^5U$ methyltransferase | rRNA (Uracil-5-)-methyltransferase RumA |
| Smp_123570 | A0A3Q0KKM5 | BHLH domain-containing protein | Aryl hydrocarbon receptor nuclear translocator |
| Smp_123590 | A0A5K4EL22 | Uncharacterized protein | Swi5-dependent recombination DNA repair protein |
| Smp_148460 | A0A3Q0KPP3 | Putative neurofibromin | GAP related domain of neurofibromin |
| Smp_162000 | A0A5K4ETP0 | UBR-type domain-containing protein | E3_UbLigase_R4 |
| Smp_246630 | A0A5K4F427 | UBC core domain-containing protein | E2 Ubiquitin conjugating enzyme |
| Smp_341570 | A0A5K4FC62 | Uncharacterized protein | Uncharacterised protein family UPF0183 |
| **Niger** | | | |
| Smp_008230 | G4V7H8 | Putative rab-18 | di-Ras2 |
| Smp_126620 | A0A3Q0KL42 | Uncharacterized protein | Ligand-binding domain of low-density lipoprotein receptor |
| Smp_165060 | A0A3Q0KRY3 | Uncharacterized protein | Uncharacterized protein |
| Smp_167890[2] | Q6BC90 | Peptide-methionine (R)-S-oxide reductase | C-terminal MsrB domain of methionine sulfoxide reductase PilB |
| Smp_313490[2] | A0A5K4F3H5 | Uncharacterized protein | Transforming protein RhoA, Rho-associated, coiled-coil |
| **Senegal** | | | |
| Smp_070780 | G4VEM1 | UDP-glucose 4-epimerase | Uridine diphosphogalactose-4-epimerase |
| Smp_123440 | A0A3Q0KKL2 | Putative fad oxidoreductase | D-amino-acid oxidase |
| Smp_123520[1] | A0A3Q0KKK4 | Putative RNA $M^5U$ methyltransferase | rRNA (Uracil-5-)-methyltransferase RumA |
| Smp_164560 | A0A3Q0KRR1 | Uncharacterized protein | Na |
| Smp_167890[2] | Q6BC90 | Peptide-methionine (R)-S-oxide reductase | C-terminal MsrB domain of methionine sulfoxide reductase PilB |
| Smp_213150 | A0A5K4EZI5 | Uncharacterized protein | Ribonucleases P/MRP protein subunit POP1 |
| Smp_313490[2] | A0A5K4F3H5 | Uncharacterized protein | Transforming protein RhoA, Rho-associated, coiled-coil |

1. Shared between the Brazil and Senegal

2. Shared between Niger and Senegal

UniProtKB descriptions are from release 2020_06

HHsearch annotations are from Le Clec'h et al. (2021)

713

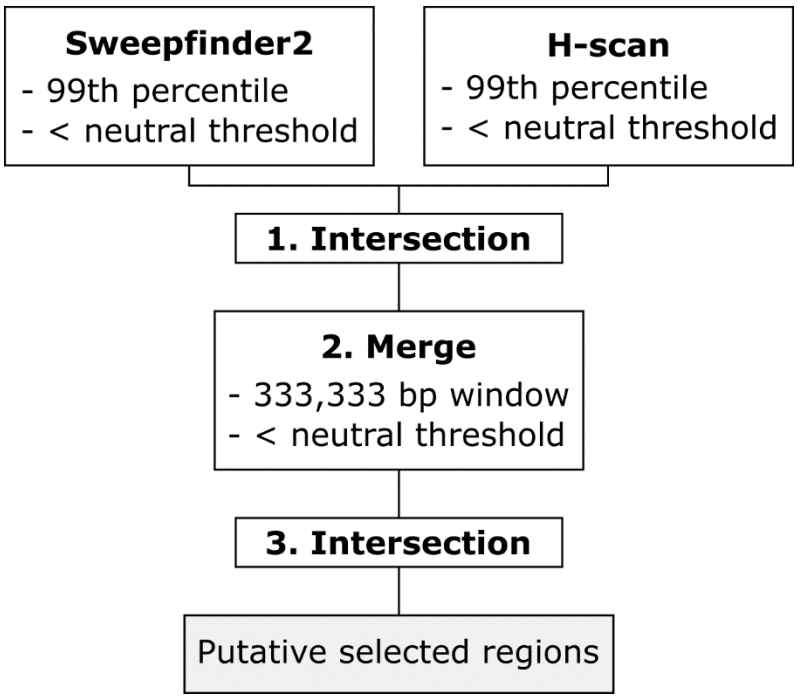
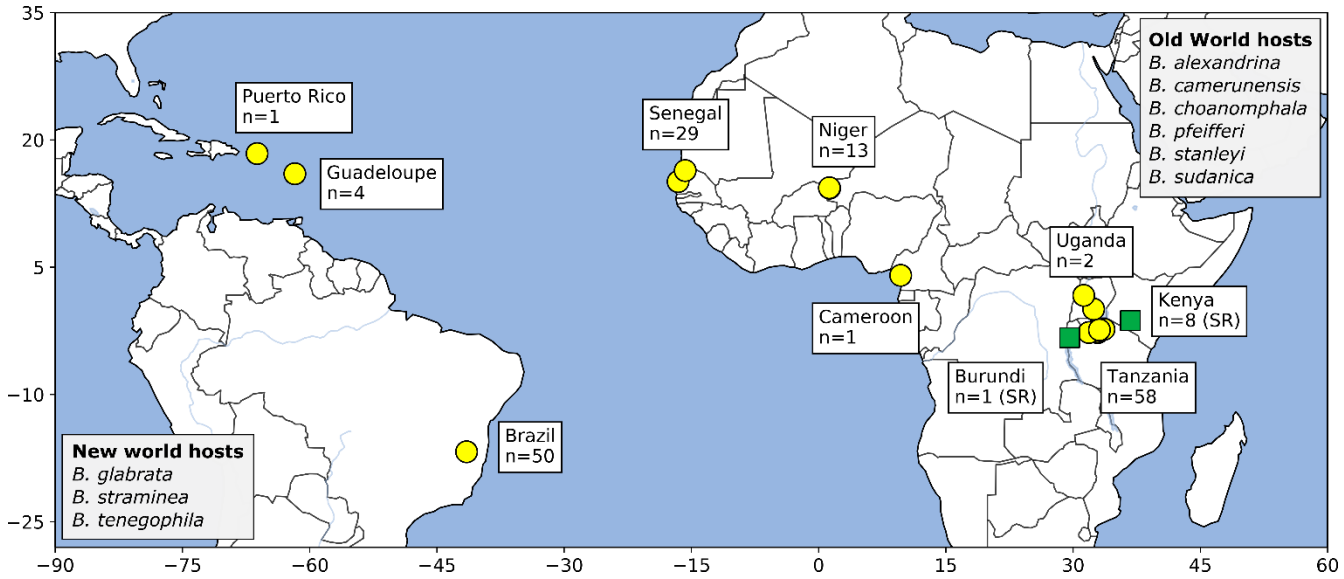



**Figure 1. Identifying regions of selection** – Flow chart shows how three separate analyses were combined to identify regions most likely experiencing positive selection. These regions are referred to as "putative selected regions". We established neutral thresholds by simulating neutrally evolving SNVs for each population. The simulated data was then run through H-Scan and Sweepfinder2 to determine the maximum expected values from neutral data for each analysis.




721

722 **Figure 2. Sampling localities** - Location and number of samples for *S. mansoni* (yellow circles) and *S.*
723 *rodhaini* (green squares) samples included in this study. Members of the genus *Biomphalaria* are the
724 predominant intermediate hosts with specific distributions of the different species involved in
725 transmission varying across Africa and S. America. Intermediate snail vectors are listed in greyed
726 boxes.

727

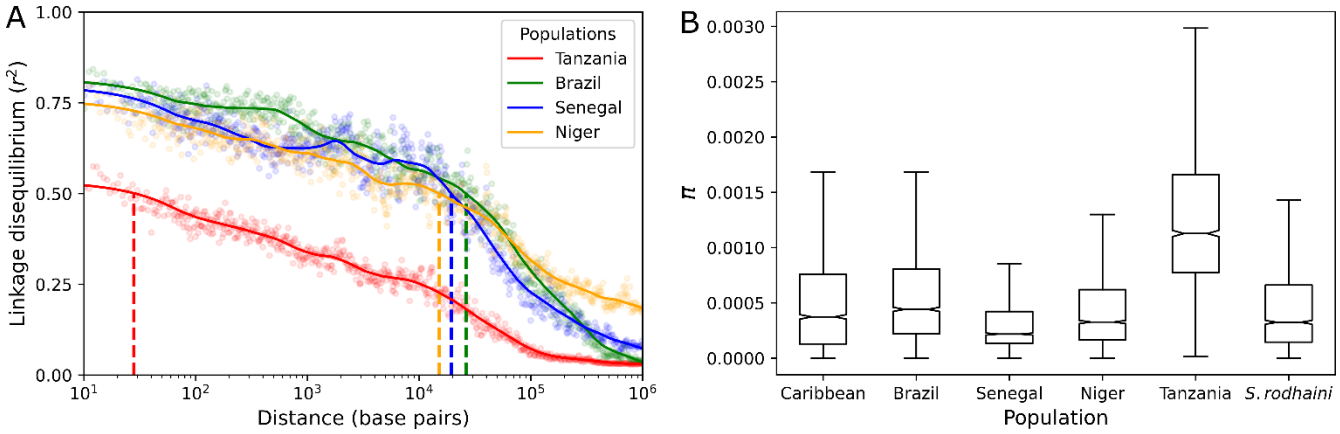
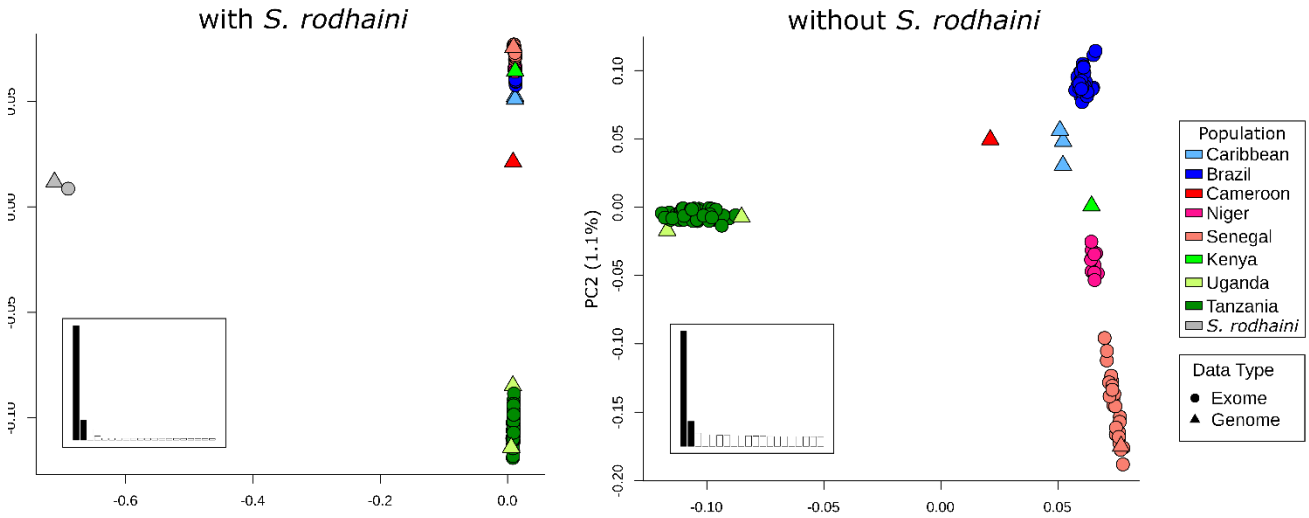



**Figure 3. Linkage Disequilibrium Decay and diversity within populations** – (A) Linkage disequilibrium between SNVs was quantified with $r^2$ values for each population. Mean $r^2$ values were taken in 500 bp windows and loess smoothed. Vertical dotted lines indicate the distance where $r^2 = 0.5$ for each population. LD decayed to $r^2$=0.5 in 28 bp (Tanzania), 15,150 bp (Niger), 19,318 bp (Senegal), and 26,196 bp (Brazil). (B) Nucleotide diversity ($\pi$) varied between *S. mansoni* populations with the highest levels of diversity occurring in E. Africa (Tanzania). Π was measured in 100 Kb windows across the autosomal chromosomes. Outliers are not shown.




737

738 **Figure 4. A PCA of unlinked autosomal SNVs** – The PCA plot that included *S. mansoni* and *S. rodhaini*
739 (left) clearly shows a large distinction between the two species with some variation within *S. mansoni*
740 along PC2. A PCA with only *S. mansoni* (right) differentiates East African *S. mansoni* along PC1. The
741 remaining *S. mansoni* samples fall along a continuum on PC2 that goes from samples in West Africa and
742 transitions to the Americas.

743

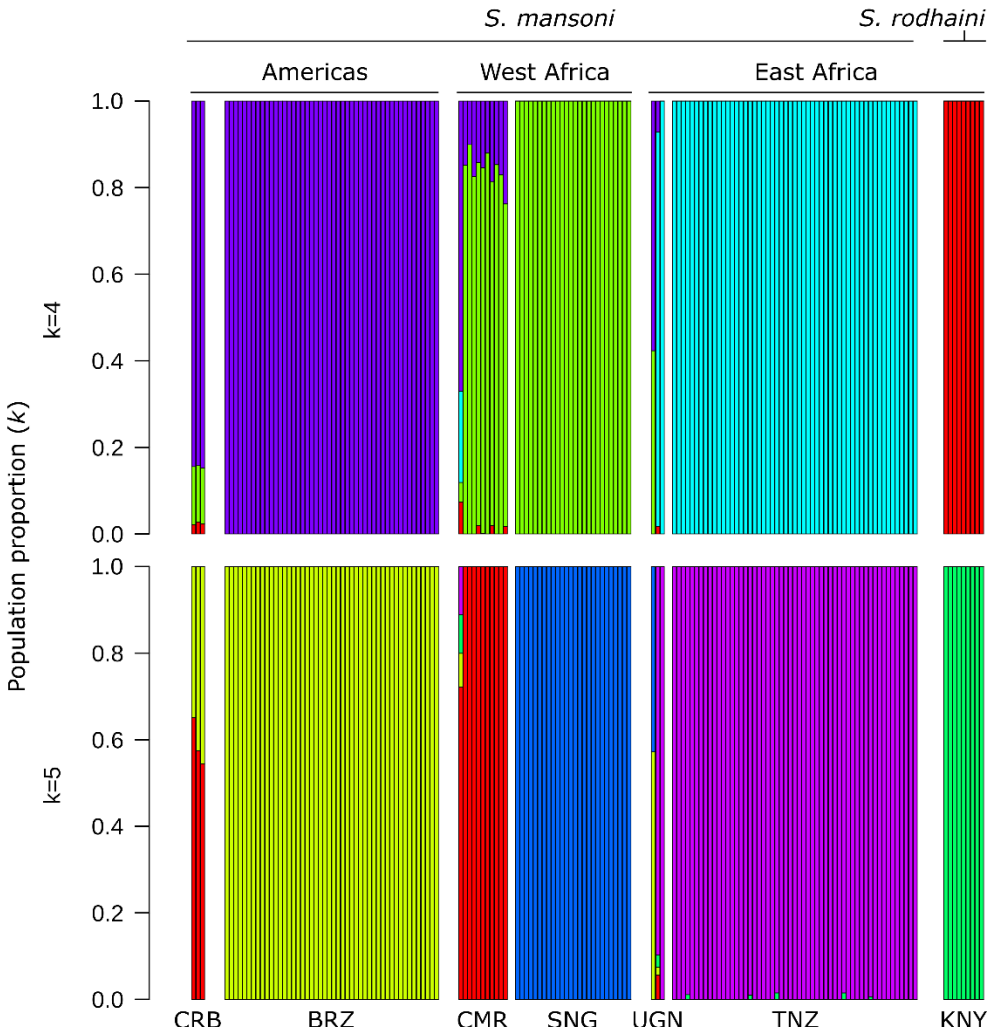



**Figure 5. Population structure in *S. mansoni*** - ADMIXTURE analyses with k=4 and k=5 populations identified clear distinctions between each of the major sampling localities (the Americas, West Africa, East Africa, and *S. rodhaini*). The population components in each of the whole genome samples from Uganda (UGN), Kenya (KNY), Cameroon (CMR), and the Caribbean (CRB) were more heterogeneous than samples with exome data. Cameroonian and Nigerien samples contain moderate proportions of the Brazilian population component. Abbreviations: "BRZ" – Brazil, "SNG" – Senegal, "NGR" – Niger, "TNZ" – Tanzania, "BRN" – Burundi.

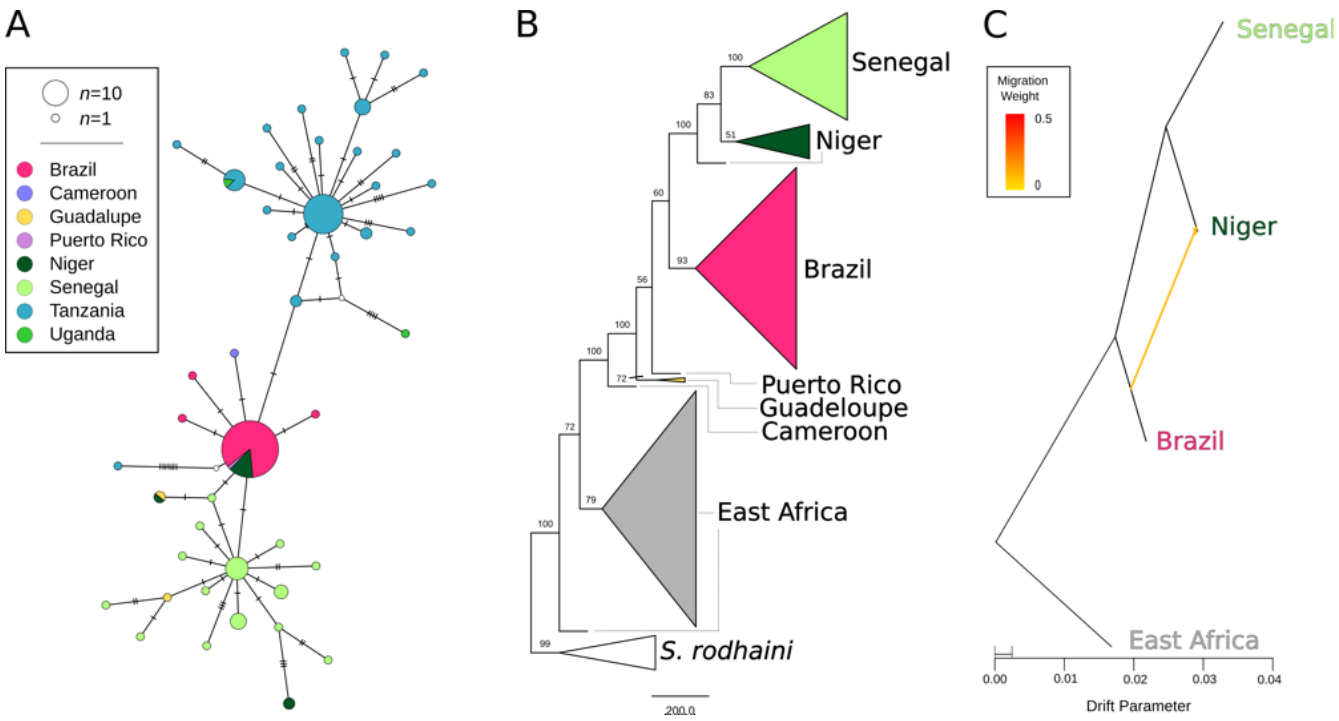



**Figure 6. Phylogenetic relationships between *S. mansoni* populations** – Multiple phylogenetic analyses and marker types were used to discern relationships between *S. mansoni* populations. (A) A median-joining haplotype network was constructed from 815 variations variants across the mitochondria of all *S. mansoni* samples. (B) A coalescent-based species tree from 100,819 parsimony informative with bootstrap values shown on each clade. Monophyletic populations are shown as a collapsed clade except in the case of E. Africa which contains samples from Tanzania and Uganda. (C) A maximum likelihood phylogenetic network of autosomal variants identified a single, weak migration edge oriented from Brazil to Niger. All three analyses identify a relationship between Senegal, Niger, and Brazil that excludes East African samples. The mitochondrial (A) and autosomal (C) networks both allow for direct relationship or allele sharing between Brazil and Niger. The species tree (B) indicates a strong relationship between Senegal and Niger that excludes Brazil (bootstrap percentage = 100).

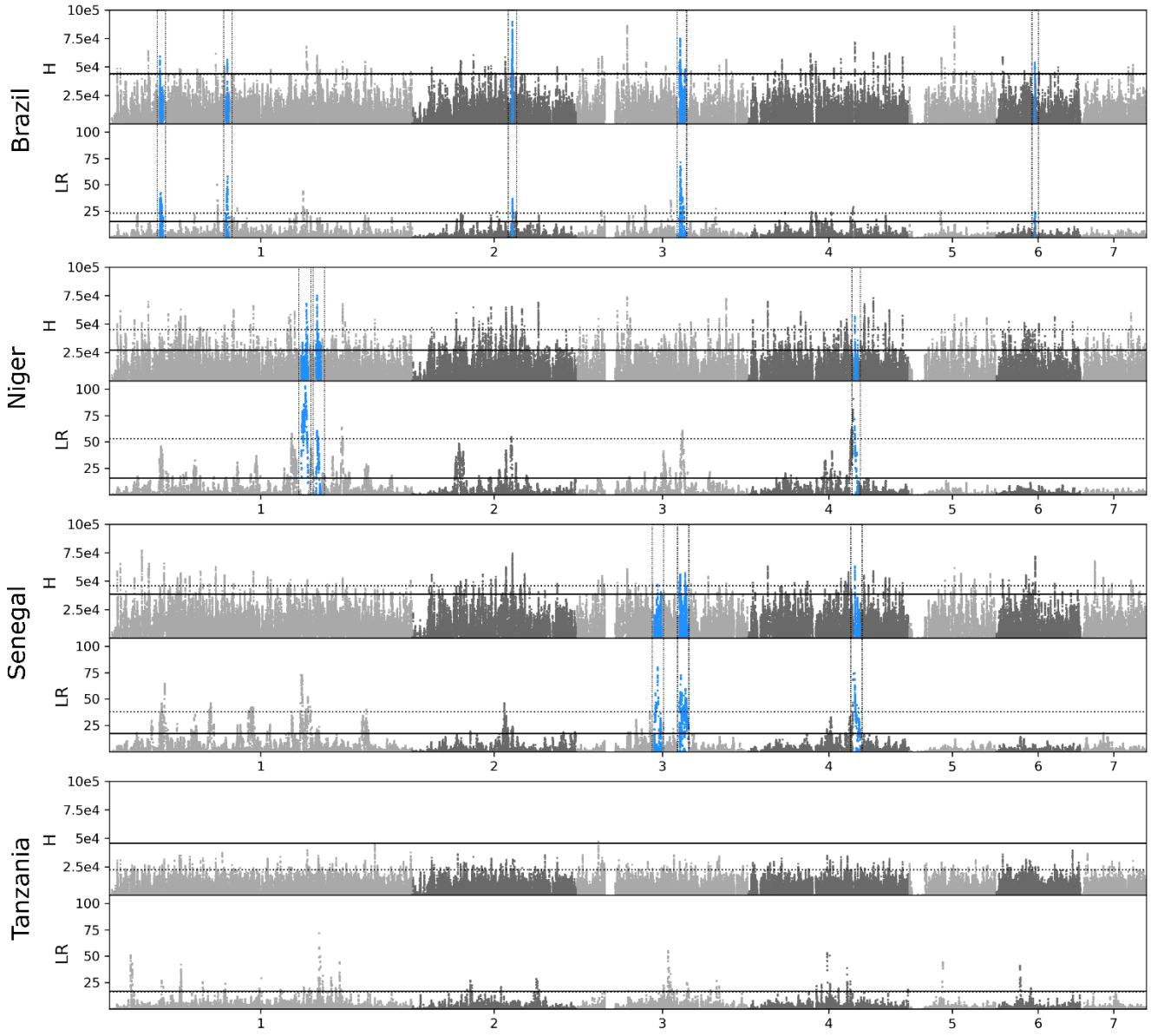



**Figure 7. Positive Selection across the *S. mansoni* genome** – Selection across the *S. mansoni* genome was calculated with haplotype (*H*) and allele frequency (likelihood ration; LR) - based methods. Dotted lines represent positions with an *H* or LR value in the 99$^{th}$ percentile. The solid black line represents the maximum *H* or LR calculated form simulated data under neutral conditions. Regions of interest (blue, boxed) were identified by finding sites where the *H* and LR values were in the >99$^{th}$ percentile and were both greater than the max *H* or LR from simulated data. Once these sites were identified we combined variants within 333,333 bp windows that showed signs of selection; an *H* or LR greater than the simulated threshold.

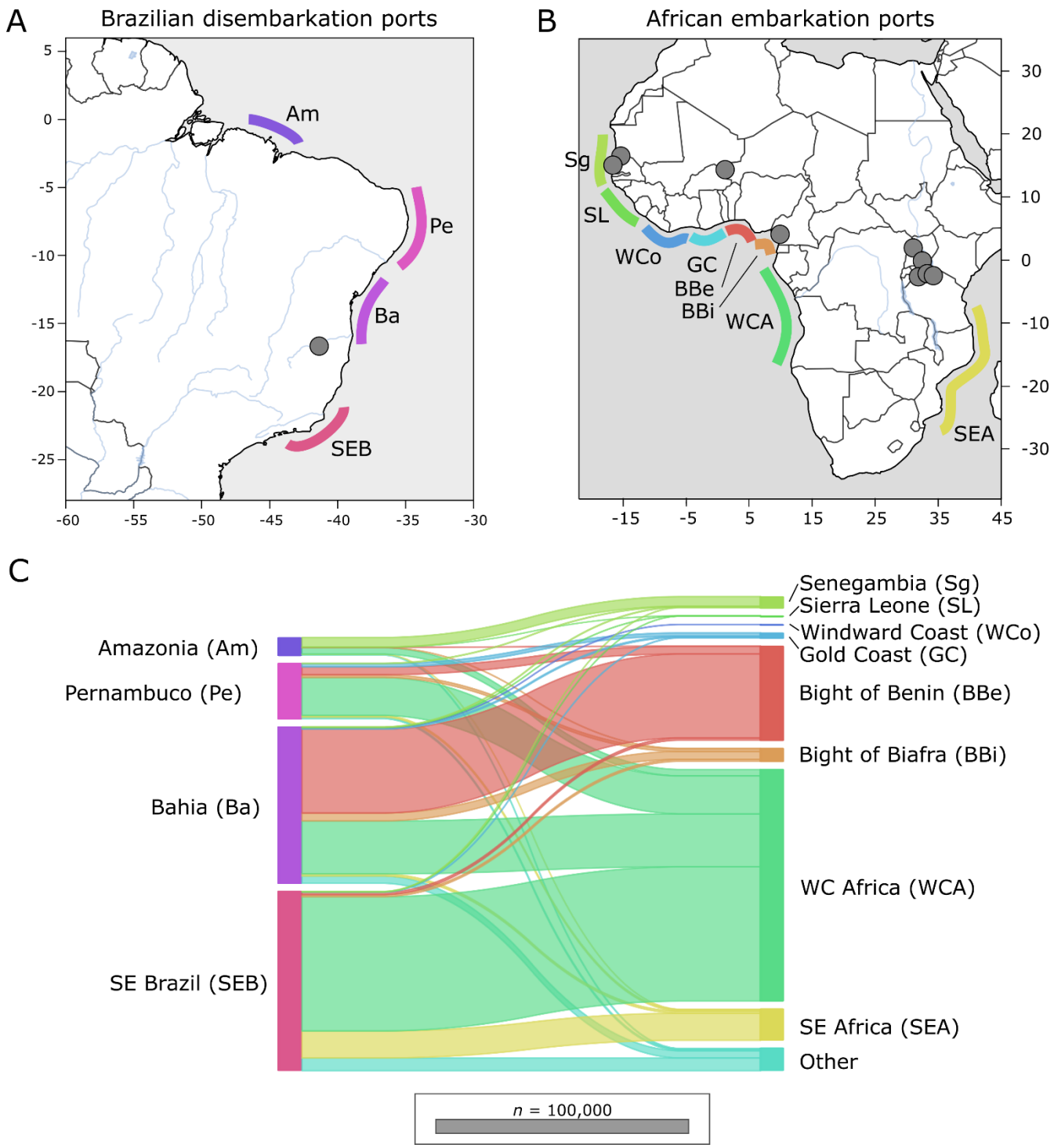



**Figure 8. Export and importation of slaves between Africa and Brazil** – During the Transatlantic Slave Trade, more than 12 million people were trafficked from Africa. Of these, 3.5 million were exported into Brazil at major ports along the east coast. Most peoples originated from regions along the West African coast from Benin to Angola. General locations of major embarkation and disembarkation ports shown along each coastline. Collection localities for *S. mansoni* are represented by grey dots. Slave exportation data is from the Voyages Database (2009).

## References


Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res, 19*(9), 1655-1664. doi:10.1101/gr.094052.109

Anderson, T. (2001). The dangers of using single locus markers in parasite epidemiology: Ascaris as a case study. *Trends Parasitol, 17*(4), 183-188. doi:10.1016/s1471-4922(00)01944-9

Anderson, T., & Enabulele, E. E. (2021). Schistosoma mansoni. *Trends Parasitol, 37*(2), 176-177. doi:10.1016/j.pt.2020.06.003

Anderson, T., Haubold, B., Williams, J. T., Estrada-Franco, J. G., Richardson, L., Mollinedo, R., . . . Day, K. P. (2000). Microsatellite markers reveal a spectrum of population structures in the malaria parasite Plasmodium falciparum. *Molecular Biology and Evolution, 17*(10), 1467-1482. doi:10.1093/oxfordjournals.molbev.a026247

Anderson, T., LoVerde, P. T., Le Clec'h, W., & Chevalier, F. D. (2018). Genetic Crosses and Linkage Mapping in Schistosome Parasites. *Trends Parasitol, 34*(11), 982-996. doi:10.1016/j.pt.2018.08.001

Ballabeni, P., & Ward, P. I. (1993). Local Adaptation of the Tremadote *Diplostomum phoxini* to the European Minnow *Phoxinus phoxinus*, its Second Intermediate Host. *Functional Ecology, 7*(1), 84-90. doi:10.2307/2389870

Bergad, L. W. (2007). *The comparative histories of slavery in Brazil, Cuba, and the United States*: Cambridge University Press New York.

Berger, D. J., Crellen, T., Lamberton, P. H., Allan, F., Tracey, A., Noonan, J. D., . . . Holroyd, N. (2021). Whole-genome sequencing of *Schistosoma mansoni* reveals extensive diversity with limited selection despite mass drug administration. *Nature Communications, 12*(1), 1-14.

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

Birky, C. W., Jr., Maruyama, T., & Fuerst, P. (1983). An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. *Genetics, 103*(3), 513-527.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114-2120. doi:10.1093/bioinformatics/btu170

Borlase, A., Rudge, J. W., Léger, E., Diouf, N. D., Fall, C. B., Diop, S. D., . . . Webster, J. P. (2021). Spillover, hybridization, and persistence in schistosome transmission dynamics at the human-animal interface. *Proc Natl Acad Sci U S A, 118*(41). doi:10.1073/pnas.2110711118

Bryant, J. E., Holmes, E. C., & Barrett, A. D. J. P. P. (2007). Out of Africa: a molecular perspective on the introduction of yellow fever virus into the Americas. *PLoS Pathog, 3*(5), e75.

Chevalier, F. D., Le Clec'h, W., Eng, N., Rugel, A. R., Assis, R. R., Oliveira, G., . . . Anderson, T. J. (2016). Independent origins of loss-of-function mutations conferring oxamniquine resistance in a Brazilian schistosome population. *International Journal for Parasitology, 46*(7), 417-424. doi:10.1016/j.ijpara.2016.03.006

Chevalier, F. D., Le Clec'h, W., McDew-White, M., Menon, V., Guzman, M. A., Holloway, S. P., . . . Anderson, T. J. C. (2019). Oxamniquine resistance alleles are widespread in Old World *Schistosoma mansoni* and predate drug deployment. *PLoS Pathog, 15*(10), e1007881. doi:10.1371/journal.ppat.1007881

Chifman, J., & Kubatko, L. (2015). Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. *Journal of Theoretical Biology, 374*, 35-47. doi:10.1016/j.jtbi.2015.03.006

Crellen, T., Allan, F., David, S., Durrant, C., Huckvale, T., Holroyd, N., . . . Cotton, J. A. (2016). Whole genome resequencing of the human parasite *Schistosoma mansoni* reveals population history and effects of selection. *Scientific reports, 6*, 20954. doi:10.1038/srep20954

Criscione, C. D., Valentim, C. L., Hirai, H., LoVerde, P. T., & Anderson, T. J. (2009). Genomic linkage map of the human blood fluke *Schistosoma mansoni*. *Genome Biology, 10*(6), R71. doi:10.1186/gb-2009-10-6-r71

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . . Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics, 27*(15), 2156-2158. doi:10.1093/bioinformatics/btr330

Davies, C. M., Webster, J. P., & Woolhous, M. E. (2001). Trade-offs in the evolution of virulence in an indirectly transmitted macroparasite. *Proc Biol Sci, 268*(1464), 251-257. doi:10.1098/rspb.2000.1367

de Oliveira, T. C., Corder, R. M., Early, A., Rodrigues, P. T., Ladeia-Andrade, S., Alves, J. M. P., . . . Ferreira, M. U. (2020). Population genomics reveals the expansion of highly inbred *Plasmodium vivax* lineages in the main malaria hotspot of Brazil. *PLoS neglected tropical diseases, 14*(10), e0008808.

DeGiorgio, M., Huber, C. D., Hubisz, M. J., Hellmann, I., & Nielsen, R. (2016). SweepFinder2: increased sensitivity, robustness and flexibility. *Bioinformatics, 32*(12), 1895-1897. doi:10.1093/bioinformatics/btw051

DeJong, R. J., Morgan, J. A., Paraense, W. L., Pointier, J.-P., Amarista, M., Ayeh-Kumi, P. F., . . . evolution. (2001). Evolutionary relationships and biogeography of *Biomphalaria* (Gastropoda: Planorbidae) with implications regarding its role as host of the human bloodfluke, *Schistosoma mansoni*. *Molecular biology, 18*(12), 2225-2239.

Dejong, R. J., Morgan, J. A., Wilson, W. D., Al-Jaser, M. H., Appleton, C. C., Coulibaly, G., . . . Loker, E. S. (2003). Phylogeography of *Biomphalaria glabrata* and *B. pfeifferi*, important intermediate hosts of *Schistosoma mansoni* in the New and Old World tropics. *Mol Ecol, 12*(11), 3041-3056. doi:10.1046/j.1365-294x.2003.01977.x

Desprès, L., Imbert-Establet, D., & Monnerot, M. (1993). Molecular characterization of mitochondrial DNA provides evidence for the recent introduction of *Schistosoma mansoni* into America. *Mol Biochem Parasitol, 60*(2), 221-229. doi:10.1016/0166-6851(93)90133-i

Doyle, S. R., Sankaranarayanan, G., Allan, F., Berger, D., Jimenez Castro, P. D., Collins, J. B., . . . Holroyd, N. (2019). Evaluation of DNA Extraction Methods on Individual Helminth Egg and Larval Stages for Whole-Genome Sequencing. *Front Genet, 10*, 826. doi:10.3389/fgene.2019.00826

Ebert, D. (1994). Virulence and local adaptation of a horizontally transmitted parasite. *Science, 265*(5175), 1084-1086. doi:10.1126/science.265.5175.1084

Eltis, D. (2001). The volume and structure of the transatlantic slave trade: A reassessment. *The William and Mary Quarterly, 58*(1), 17-46.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular ecology, 14*(8), 2611-2620.

Files, V. S. (1951). A study of the vector-parasite relationships in *Schistosoma mansoni*. *Parasitology, 41*(3-4), 264-269.

Fletcher, M., LoVerde, P. T., & Woodruff, D. S. (1981). Genetic variation in *Schistosoma mansoni*: enzyme polymorphisms in populations from Africa, Southwest Asia, South America, and the West Indies. *American Journal of Tropical Medicine and Hygiene, 30*(2), 406-421. doi:10.4269/ajtmh.1981.30.406

Forni, D., Pontremoli, C., Clerici, M., Pozzoli, U., Cagliani, R., & Sironi, M. (2020). Recent Out-of-Africa Migration of Human Herpes Simplex Viruses. *Molecular Biology and Evolution, 37*(5), 1259-1271. doi:10.1093/molbev/msaa001

Fulford, A. J., Butterworth, A. E., Ouma, J. H., & Sturrock, R. F. (1995). A statistical approach to schistosome population dynamics and estimation of the life-span of Schistosoma mansoni in man. *Parasitology, 110 ( Pt 3)*, 307-316. doi:10.1017/s0031182000080896

Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., . . . Sandhu, M. S. (2015). The African Genome Variation Project shapes medical genetics in Africa. *Nature, 517*(7534), 327-332. doi:10.1038/nature13997

Guzmán-Solís, A. A., Villa-Islas, V., Bravo-López, M. J., Sandoval-Velasco, M., Wesp, J. K., Gómez-Valdés, J. A., . . . Ávila Arcos, M. C. (2021). Ancient viral genomes reveal introduction of human pathogenic viruses into Mexico during the transatlantic slave trade. *Elife, 10*. doi:10.7554/eLife.68612

Hahn, M. W., & Hibbins, M. S. (2019). A Three-Sample Test for Introgression. *Molecular Biology and Evolution, 36*(12), 2878-2882. doi:10.1093/molbev/msz178

Hailegebriel, T., Nibret, E., & Munshea, A. (2020). Prevalence of *Schistosoma mansoni* and *S. haematobium* in Snail Intermediate Hosts in Africa: A Systematic Review and Meta-analysis. *Journal of tropical medicine, 2020*.

International Helminth Genomes Consortium. (2019). Comparative genomics of the major parasitic worms. *Nat Genet, 51*(1), 163-174. doi:10.1038/s41588-018-0262-1

Jorgenson, A., Kristensen, T. K., & Stothard, J. R. (2007). Phylogeny and biogeography of African Biomphalaria (Gastropoda: Planorbidae), with emphasis on endemic species of the great East African lakes. *Zoological Journal of the Linnean Society, 151*(2), 337-349.

Kabongo, M. M., Linsuke, S., Baloji, S., Mukunda, F., Raquel, I. D. L., Stauber, C., . . . Lutumba, P. (2018). Schistosoma mansoni infection and its association with nutrition and health outcomes: a household survey in school-aged children living in Kasansa, Democratic Republic of the Congo. *Pan African Medical Journal 31*, 197. doi:10.11604/pamj.2018.31.197.16364

Kelleher, J., Etheridge, A. M., & McVean, G. (2016). Efficient Coalescent Simulation and Genealogical Analysis for Large Sample Sizes. *PLoS Computational Biology, 12*(5), e1004842. doi:10.1371/journal.pcbi.1004842

Kengne-Fokam, A. C., Nana-Djeunga, H. C., Bagayan, M., & Njiokou, F. (2018). *Biomphalaria camerunensis* as a viable alternative intermediate host for *Schistosoma mansoni* in southern Cameroon. *Parasites & vectors, 11*(1), 1-9.

Köster, J., & Rahmann, S. (2012). Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics, 28*(19), 2520-2522.

Köster, J., & Rahmann, S. (2018). Snakemake-a scalable bioinformatics workflow engine. *Bioinformatics, 34*(20), 3600. doi:10.1093/bioinformatics/bty350

Le Clec'h, W., Chevalier, F. D., McDew-White, M., Allan, F., Webster, B. L., Gouvras, A. N., . . . Anderson, T. J. C. (2018). Whole genome amplification and exome sequencing of archived schistosome miracidia. *Parasitology, 145*(13), 1739-1747. doi:10.1017/S0031182018000811

Le Clec'h, W., Chevalier, F. D., McDew-White, M., Menon, V., Arya, G. A., & Anderson, T. J. C. (2021). Genetic architecture of transmission stage production and virulence in schistosome parasites. *Virulence, 12*(1), 1508-1526. doi:10.1080/21505594.2021.1932183

Le Clec'h, W., Chevalier, F. D., Mattos, A. C. A., Strickland, A., Diaz, R., McDew-White, M., . . . Anderson, T. J. C. (2021). Genetic analysis of praziquantel response in schistosome parasites implicates a Transient Receptor Potential channel. 2021.2006.2009.447779. doi:10.1101/2021.06.09.447779 %J bioRxiv

Leblois, R., Kuhls, K., François, O., Schönian, G., & Wirth, T. (2011). Guns, germs and dogs: On the origin of *Leishmania chagasi*. *Infection, Genetics and Evolution, 11*(5), 1091-1095. doi:10.1016/j.meegid.2011.04.004

Léger, E., Borlase, A., Fall, C. B., Diouf, N. D., Diop, S. D., Yasenev, L., . . . Webster, J. P. (2020). Prevalence and distribution of schistosomiasis in human, livestock, and snail populations in northern Senegal: a One Health epidemiological study of a multi-host system. *Lancet Planet Health, 4*(8), e330-e342. doi:10.1016/s2542-5196(20)30129-7

Leger, E., & Webster, J. P. (2017). Hybridizations within the genus *Schistosoma*: implications for evolution, epidemiology and control. *Parasitology, 144*(1), 65-80.

Leigh, J. W., & Bryant, D. (2015). POPART: full-feature software for haplotype network construction. *Methods in Ecology Evolution, 6*(9), 1110-1116.

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics, 26*(5), 589-595. doi:10.1093/bioinformatics/btp698

Lively, C. M. (1989). Adaptation by a parasitic trematode to local populations of its snail host. *Evolution, 43*(8), 1663-1671. doi:10.1111/j.1558-5646.1989.tb02616.x

Luu, K., Bazin, E., & Blum, M. G. (2017). pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular ecology resources, 17*(1), 67-77.

Mackay, T. F. C., & Huang, W. (2018). Charting the genotype-phenotype map: lessons from the Drosophila melanogaster Genetic Reference Panel. *Wiley Interdiscip Rev Dev Biol, 7*(1). doi:10.1002/wdev.289

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., . . . DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res, 20*(9), 1297-1303. doi:10.1101/gr.107524.110

Miles, A., Ralph, P., Rae, S., & Pisupati, R. (2019). scikit-allel (Version 1.2.1).

Mitta, G., Gourbal, B., Grunau, C., Knight, M., Bridger, J., & Théron, A. (2017). The compatibility between *Biomphalaria glabrata* snails and *Schistosoma mansoni:* an increasingly complex puzzle. In *Advances In Parasitology* (Vol. 97, pp. 111-145): Elsevier.

Morand, S., Manning, S. D., & Woolhouse, M. E. (1996). Parasite-host coevolution and geographic patterns of parasite infectivity and host susceptibility. *Proc Biol Sci, 263*(1366), 119-128. doi:10.1098/rspb.1996.0019

Morgan, J., DeJong, R. J., Lwambo, N. J., Mungai, B. N., Mkoji, G. M., & Loker, E. S. (2003). First report of a natural hybrid between *Schistosoma mansoni* and *S. rodhaini*. *Journal of Parasitology, 89*(2), 416-418.

Morgan, J. A., Dejong, R. J., Adeoye, G. O., Ansa, E. D., Barbosa, C. S., Bremond, P., . . . Loker, E. S. (2005). Origin and diversification of the human parasite *Schistosoma mansoni*. *Molecular ecology, 14*(12), 3889-3902. doi:10.1111/j.1365-294X.2005.02709.x

Mutuku, M. W., Laidemitt, M. R., Spaan, J. M., Mwangi, I. N., Ochanda, H., Steinauer, M. L., . . . Mkoji, G. M. (2021). Comparative Vectorial Competence of Biomphalaria sudanica and Biomphalaria choanomphala, Snail Hosts of Schistosoma mansoni, From Transmission Hotspots In Lake Victoria, Western Kenya. *J Parasitol, 107*(2), 349-357. doi:10.1645/20-138

Mutuku, M. W., Lu, L., Otiato, F. O., Mwangi, I. N., Kinuthia, J. M., Maina, G. M., . . . Mkoji, G. M. (2017). A Comparison of Kenyan Biomphalaria pfeifferi and B. Sudanica as Vectors for *Schistosoma mansoni*, Including a Discussion of the Need to Better Understand the Effects of Snail Breeding Systems on Transmission. *J Parasitol, 103*(6), 669-676. doi:10.1645/17-72

Nadeau, S. A., Vaughan, T. G., Scire, J., Huisman, J. S., & Stadler, T. (2021). The origin and early spread of SARS-CoV-2 in Europe. *Proceedings of the National Academy of Sciences, 118*(9).

Neafsey, D. E., Schaffner, S. F., Volkman, S. K., Park, D., Montgomery, P., Milner, D. A., Jr., . . . Wirth, D. F. (2008). Genome-wide SNP genotyping highlights the role of natural selection in Plasmodium falciparum population divergence. *Genome Biology, 9*(12), R171. doi:10.1186/gb-2008-9-12-r171

Nychka, D., Furrer, R., Paige, J., & Sain, S. (2017). fields: Tools for Spatial Data (Version 11.6): University Corporation for Atmospheric Research. Retrieved from https://github.com/NCAR/Fields

Ortiz, E. M. (2019). vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis. Retrieved from DOI:10.5281/zenodo.2540861

978 Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., . . . Reich, D. (2012). Ancient
979 admixture in human history. *Genetics, 192*(3), 1065-1093. doi:10.1534/genetics.112.145037
980 Pedersen, B. S., & Quinlan, A. R. (2018). Mosdepth: quick coverage calculation for genomes and
981 exomes. *Bioinformatics, 34*(5), 867-868. doi:10.1093/bioinformatics/btx699
982 Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide
983 allele frequency data. *PLoS Genetics, 8*(11), e1002967. doi:10.1371/journal.pgen.1002967
984 Platt, R. N., McDew-White, M., Le Clec'h, W., Chevalier, F. D., Allan, F., Emery, A. M., . . . Anderson,
985 T. J. C. (2019). Ancient hybridization and adaptive introgression of an invadolysin gene in
986 schistosome parasites. *Molecular Biology and Evolution, 36*(10), 2127-2142.
987 doi:10.1093/molbev/msz154
988 Portet, A., Pinaud, S., Chaparro, C., Galinier, R., Dheilly, N. M., Portela, J., . . . Gourbal, B. (2019).
989 Sympatric versus allopatric evolutionary contexts shape differential immune response in
990 Biomphalaria / Schistosoma interaction. *PLoS Pathog, 15*(3), e1007647.
991 doi:10.1371/journal.ppat.1007647
992 Protasio, A. V., Tsai, I. J., Babbage, A., Nichol, S., Hunt, M., Aslett, M. A., . . . Berriman, M. (2012). A
993 systematically improved high quality genome and transcriptome of the human blood fluke
994 Schistosoma mansoni. *PLoS Negl Trop Dis, 6*(1), e1455. doi:10.1371/journal.pntd.0001455
995 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., . . . Sham, P. C.
996 (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses.
997 *American Journal of Human Genetics, 81*(3), 559-575. doi:10.1086/519795
998 Rey, O., Toulza, E., Chaparro, C., Allienne, J. F., Kincaid-Smith, J., Mathieu-Begné, E., . . . Boissier, J.
999 (2021). Diverging patterns of introgression from *Schistosoma bovis* across *S. haematobium*
1000 African lineages. *PLoS Pathog, 17*(2), e1009313. doi:10.1371/journal.ppat.1009313
1001 Rey, O., Webster, B. L., Huyse, T., Rollinson, D., Van den Broeck, F., Kincaid-Smith, J., . . . Boissier, J.
1002 (2021). Population genetics of African *Schistosoma* species. *Infection, Genetics and Evolution,*
1003 *89*, 104727. doi:10.1016/j.meegid.2021.104727
1004 Rius, M., Bourne, S., Hornsby, H. G., & Chapman, M. A. (2015). Applications of next-generation
1005 sequencing to the study of biological invasions. *Current Zoology, 61*(3), 488-504.
1006 Rodrigues, P. T., Valdivia, H. O., de Oliveira, T. C., Alves, J. M. P., Duarte, A. M. R., Cerutti-Junior,
1007 C., . . . Hirano, Z. M. (2018). Human migration and the spread of malaria parasites to the New
1008 World. *Scientific reports, 8*(1), 1-13.
1009 Sanneh, B., Joof, E., Sanyang, A. M., Renneker, K., Camara, Y., Sey, A. P., . . . Ogoussan, K. (2017).
1010 Field evaluation of a schistosome circulating cathodic antigen rapid test kit at point-of-care for
1011 mapping of schistosomiasis endemic districts in The Gambia. *PLoS One, 12*(8), e0182003.
1012 doi:10.1371/journal.pone.0182003
1013 Schlamp, F., van der Made, J., Stambler, R., Chesebrough, L., Boyko, A. R., & Messer, P. W. (2016).
1014 Evaluating the performance of selection scans to detect selective sweeps in domestic dogs.
1015 *Molecular ecology, 25*(1), 342-356. doi:10.1111/mec.13485
1016 Schwabl, P., Boité, M. C., Bussotti, G., Jacobs, A., Andersson, B., Moreira, O., . . . Traub-Csekö, Y.
1017 (2021). Colonization and genetic diversification processes of *Leishmania infantum* in the
1018 Americas. *Communications biology, 4*(1), 1-13.
1019 Sherpa, S., & Després, L. (2021). The evolutionary dynamics of biological invasions: A multi-approach
1020 perspective. *Evolutionary Applications*.
1021 Shortt, J. A., Card, D. C., Schield, D. R., Liu, Y., Zhong, B., Castoe, T. A., . . . Pollock, D. D. (2017).
1022 Whole Genome Amplification and Reduced-Representation Genome Sequencing of
1023 Schistosoma japonicum Miracidia. *PLoS neglected tropical diseases, 11*(1), e0005292.
1024 doi:10.1371/journal.pntd.0005292
1025 Slave Voyages Database. (2009). Voyages: The Trans-Atlantic Slave Trade Database. Retrieved from
1026 https://www.slavevoyages.org

Small, S. T., Labbe, F., Coulibaly, Y. I., Nutman, T. B., King, C. L., Serre, D., & Zimmerman, P. A. (2019). Human Migration and the Spread of the Nematode Parasite *Wuchereria bancrofti*. *Molecular Biology and Evolution, 36*(9), 1931-1941. doi:10.1093/molbev/msz116

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

Steinauer, M., Mwangi, I. N., Maina, G. M., Kinuthia, J. M., Mutuku, M. W., Agola, E. L., . . . Loker, E. S. (2008). Interactions between natural populations of human and rodent schistosomes in the Lake Victoria region of Kenya: a molecular epidemiological approach. *PLoS neglected tropical diseases, 2*(4), e222.

Steinegger, M., Meier, M., Mirdita, M., Vöhringer, H., Haunsberger, S. J., & Söding, J. (2019). HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics, 20*(1), 473. doi:10.1186/s12859-019-3019-7

Steverding, D. (2020). The spreading of parasites by human migratory activities. *Virulence, 11*(1), 1177-1191. doi:10.1080/21505594.2020.1809963

Stothard, J. R., Kayuni, S. A., Al-Harbi, M. H., Musaya, J., & Webster, B. L. (2020). Future schistosome hybridizations: Will all *Schistosoma haematobium* hybrids please stand-up! *PLoS neglected tropical diseases, 14*(7), e0008201. doi:10.1371/journal.pntd.0008201

Swofford, D. L. (2003). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods) Sunderland, Massachusetts.: Sinauer Associates. Retrieved from https://paup.phylosolutions.com/

Tennessen, J. A., Bollmann, S. R., Peremyslova, E., Kronmiller, B. A., Sergi, C., Hamali, B., & Blouin, M. S. (2020). Clusters of polymorphic transmembrane genes control resistance to schistosomes in snail vectors. *Elife, 9*, e59395. doi:10.7554/eLife.59395

Tennessen, J. A., Theron, A., Marine, M., Yeh, J. Y., Rognon, A., & Blouin, M. S. (2015). Hyperdiverse gene cluster in snail host conveys resistance to human schistosome parasites. *PLoS Genetics, 11*(3), e1005067. doi:10.1371/journal.pgen.1005067

Théron, A. (1989). Hybrids between *Schistosoma mansoni* and *S. rodhaini*: characterization by cercarial emergence rhythms. *Parasitology, 99 Pt 2*, 225-228. doi:10.1017/s0031182000058674

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. *Infection, Genetics and Evolution, 26*, 80-88. doi:10.1016/j.meegid.2014.04.025

Van den Broeck, F., Meurs, L., Raeymaekers, J., Boon, N., Dieye, T., Volckaert, F., . . . Huyse, T. (2014). Inbreeding within human *Schistosoma mansoni*: do host-specific factors shape the genetic composition of parasite populations? *Heredity, 113*(1), 32-41.

Vidigal, T., Kissinger, J., Caldeira, R. L., Pires, E., Monteiro, E., Simpson, A., & Carvalho, O. (2000). Phylogenetic relationships among Brazilian *Biomphalaria* species (Mollusca: Planorbidae) based upon analysis of ribosomal ITS2 sequences. *Parasitology, 121*(6), 611-620.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., . . . Bright, J. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature methods, 17*(3), 261-272.

Webster, B. L., Webster, J. P., Gouvras, A. N., Garba, A., Lamine, M. S., Diaw, O. T., . . . Mubila, L. (2013). DNA 'barcoding'of *Schistosoma mansoni* across sub-Saharan Africa supports substantial within locality diversity and geographical separation of genotypes. *Acta tropica, 128*(2), 250-260.

Webster, J. P., Davies, C. M., Hoffman, J. I., Ndamba, J., Noble, L. R., & Woolhouse, M. E. (2001). Population genetics of the schistosome intermediate host Biomphalaria pfeifferi in the Zimbabwean highveld: implications for co-evolutionary theory. *Ann Trop Med Parasitol, 95*(2), 203-214. doi:10.1080/00034980120041062

Webster, J. P., Gower, C. M., & Blair, L. (2004). Do hosts and parasites coevolve? Empirical support from the *Schistosoma* system. *Am Nat, 164 Suppl 5*, S33-51. doi:10.1086/424607

Webster, J. P., & Woolhouse, M. E. J. (1998). Selection and strain specificity of compatibility between snail intermediate hosts and their parasitic schistosomes. *Evolution, 52*(6), 1627-1634. doi:10.1111/j.1558-5646.1998.tb02243.x

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution, 38*(6), 1358-1370. doi:10.1111/j.1558-5646.1984.tb05657.x

Woolhouse, M. E., & Webster, J. P. (2000). In search of the red queen. *Parasitol Today, 16*(12), 506-508. doi:10.1016/s0169-4758(00)01820-2

Yalcindag, E., Elguero, E., Arnathau, C., Durand, P., Akiana, J., Anderson, T. J., . . . Prugnolle, F. (2012). Multiple independent introductions of *Plasmodium falciparum* in South America. *Proceedings of the National Academy of Sciences, 109*(2), 511-516. doi:10.1073/pnas.1119058109