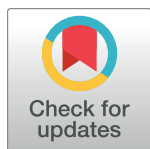RESEARCH ARTICLE

Oxamniquine resistance alleles are widespread in Old World *Schistosoma mansoni* and predate drug deployment

Frédéric D. Chevalier[1]*, **Winka Le Clec'h**[1], **Marina McDew-White**[1], **Vinay Menon**[1], **Meghan A. Guzman**[2], **Stephen P. Holloway**[3], **Xiaohang Cao**[3], **Alexander B. Taylor**[3,4], **Safari Kinung'hi**[5], **Anouk N. Gouvras**[6,7], **Bonnie L. Webster**[6,7], **Joanne P. Webster**[6,8], **Aidan M. Emery**[6,7], **David Rollinson**[6,7], **Amadou Garba Djirmay**[9,10], **Khalid M. Al Mashikhi**[11], **Salem Al Yafae**[11], **Mohamed A. Idris**[12], **Hélène Moné**[13], **Gabriel Mouahid**[13], **P. John Hart**[3,4], **Philip T. LoVerde**[2], **Timothy J. C. Anderson**[1]*

1 Texas Biomedical Research Institute, San Antonio, Texas, United States of America, 2 Departments of Pathology and University of Texas Health Science Center at San Antonio, San Antonio, Texas, United States of America, 3 Biochemistry & Structural Biology, University of Texas Health Science Center at San Antonio, San Antonio, Texas, United States of America, 4 X-ray Crystallography Core Laboratory, University of Texas Health Science Center at San Antonio, San Antonio, Texas, United States of America, 5 National Institute for Medical Research, Mwanza, United Republic of Tanzania, 6 London Centre for Neglected Tropical Disease Research (LCNDTR), Imperial Collge, London, United Kingdom, 7 Wolfson Wellcome Biomedical Laboratories, Natural History Museum, London, United Kingdom, 8 Centre for Emerging, Endemic and Exotic Diseases (CEEED), Royal Veterinary College, University of London, United Kingdom, 9 Réseau International Schistosomiases Environnemental Aménagement et Lutte (RISEAL), Niamey, Niger, 10 World Health Organization, Geneva, Switzerland, 11 Directorate General of Health Services, Dhofar Governorate, Salalah, Sultanate of Oman, 12 Sultan Qaboos University, Muscat, Sultanate of Oman, 13 Host-Pathogen-Environment Interactions laboratory, University of Perpignan, Perpignan, France

* fcheval@txbiomed.org (FDC); tanderso@txbiomed.org (TJCA)

Abstract

Do mutations required for adaptation occur *de novo*, or are they segregating within populations as standing genetic variation? This question is key to understanding adaptive change in nature, and has important practical consequences for the evolution of drug resistance. We provide evidence that alleles conferring resistance to oxamniquine (OXA), an antischistosomal drug, are widespread in natural parasite populations under minimal drug pressure and predate OXA deployment. OXA has been used since the 1970s to treat *Schistosoma mansoni* infections in the New World where *S. mansoni* established during the slave trade. Recessive loss-of-function mutations within a parasite sulfotransferase (SmSULT-OR) underlie resistance, and several verified resistance mutations, including a deletion (p. E142del), have been identified in the New World. Here we investigate sequence variation in *SmSULT-OR* in *S. mansoni* from the Old World, where OXA has seen minimal usage. We sequenced exomes of 204 *S. mansoni* parasites from West Africa, East Africa and the Middle East, and scored variants in *SmSULT-OR* and flanking regions. We identified 39 non-synonymous SNPs, 4 deletions, 1 duplication and 1 premature stop codon in the *SmSULT-OR* coding sequence, including one confirmed resistance deletion (p.E142del). We expressed recombinant proteins and used an *in vitro* OXA activation assay to functionally validate the OXA-resistance phenotype for four predicted OXA-resistance mutations. Three aspects of the data are of particular interest: (i) segregating OXA-resistance alleles are
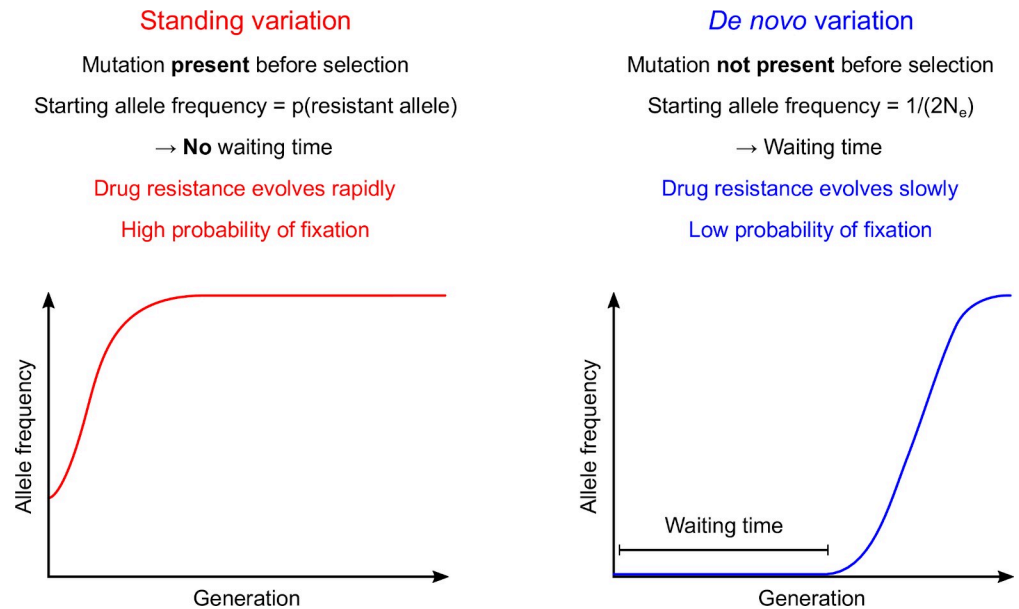
widespread in Old World populations (4.29–14.91% frequency), despite minimal OXA usage, (ii) two OXA-resistance mutations (p.W120R, p.N171IfsX28) are particularly common (>5%) in East African and Middle-Eastern populations, (iii) the p.E142del allele has identical flanking SNPs in both West Africa and Puerto Rico, suggesting that parasites bearing this allele colonized the New World during the slave trade and therefore predate OXA deployment. We conclude that standing variation for OXA resistance is widespread in *S. mansoni*.

Author summary

It has been argued that drug resistance is unlikely to spread rapidly in helminth parasites infecting humans. This is based, at least in part, on the premise that resistance mutations are rare or absent within populations prior to treatment, and take a long time to reach appreciable frequencies because helminth parasite generation time is long. This argument is critically dependent on the starting frequency of resistance alleles–if high levels of "standing variation" for resistance are present prior to deployment of treatment, resistance may spread rapidly. We examined frequencies of oxamniquine resistance alleles present in *Schistosoma mansoni* from Africa and the Middle East where oxamniquine has seen minimal use. We found that oxamniquine resistance alleles are widespread in the Old World, ranging from 4.29% in the Middle East to 14.91% in East African parasite populations. Furthermore, we show that resistance alleles from West African and the Caribbean schistosomes share a common origin, suggesting that these alleles travelled to the New World with *S. mansoni* during the transatlantic slave trade. Together, these results demonstrate extensive standing variation for oxamniquine resistance. Our results have important implications for both drug treatment policies and drug development efforts, and demonstrate the power of molecular surveillance approaches for guiding helminth control.

Introduction

The rate at which drug resistance alleles (or any other beneficial alleles) spread within populations in response to a drug treatment (or any other selection pressure) is critically dependent on the starting frequency of resistance alleles in the population when new drugs are deployed [1,2]. If no such alleles are present when a novel drug is introduced, then there is a waiting time for resistance alleles to arise. Furthermore, because the starting allele frequency will be $1/(2N_e)$, where N_e is the effective population size, the vast majority of resistance alleles that arise will be lost due to genetic drift and fail to establish [2,3] (Fig 1). The barrier to establishment is particularly severe for recessive traits because these will be present in heterozygotes at low frequency and therefore not exposed to selection. This effect, otherwise known as "Haldane's sieve" [4], may be particularly relevant for drug resistance evolution in diploid pathogens, because resistance mutations in drug targets typically result in phenotypic resistance only when two copies are present. However, if resistance alleles are already segregating as standing variation in pathogen populations when drug treatment is initiated, then there is no waiting time for mutations to arise. Furthermore, there is a high probability of fixation and spread is rapid because many resistance alleles are present in homozygous recessive genotypes and therefore exposed to selection. Hence, whether resistance alleles (or alleles for other traits)

Standing variation

Mutation **present** before selection

Starting allele frequency = p(resistant allele)

→ **No** waiting time

Drug resistance evolves rapidly

High probability of fixation

De novo variation

Mutation **not present** before selection

Starting allele frequency = $1/(2N_e)$

→ Waiting time

Drug resistance evolves slowly

Low probability of fixation

Fig 1. Drug resistance evolution from standing variation or *de novo* mutation. Drug resistance alleles spread rapidly and reach fixation with a high probability when drug resistance alleles are already present as standing variation. In contrast, if resistance alleles are absent when treatment is initiated, resistance mutation must arise *de novo*, so there is a waiting time before the resistance alleles appear, and most resistance alleles are lost by genetic drift, so the probability of establishment and fixation is low.

https://doi.org/10.1371/journal.ppat.1007881.g001

arise *de novo* or are segregating as standing variation, is a central question in evolutionary biology and public health, and key to predicting the effective "shelf life" of new drug treatments [3,5].

There is growing evidence for resistance to anti-helminthics in several helminths infecting humans, including *Onchocerca volvulus* [6], the filarial nematode causing river blindness, and in soil transmitted helminths (hookworm, whipworm and *Ascaris* roundworms) [7] which cumulatively infect over one billion people worldwide [8]. However, perhaps the best understood is oxamniquine resistance in schistosome blood flukes, for which the mechanism of drug action and the genetic and molecular basis for drug resistance are now known. Oxamniquine (OXA) kills *Schistosoma mansoni*, but is ineffective against two other major schistosome species infecting humans (*S. haematobium* and *S. japonicum*) [9]. OXA is a pro-drug that is activated by a schistosome sulfotransferase (SmSULT-OR) encoded on chromosome 6 [10]. Loss-of-function mutations within SmSULT-OR result in resistance: only parasites that are homozygous for resistance alleles are phenotypically resistant (OXA-R). We initially identified the locus and mutation, a single amino acid deletion (p.E142del), underlying OXA-R using a genetic cross involving a resistant parasite line selected from a parasite isolate from a Puerto Rican patient in 1971 [11]. Additionally we identified a second mutation (p.C35R) in *S. mansoni* obtained from an incurable Brazilian patient [12]. These two mutations result in disruption of the active site of the SmSULT-OR and inability to activate OXA [10]. Both p.E142del and p.C35R, and two additional mutations resulting in truncated proteins, were subsequently identified in a field survey of SmSULT-OR genetic variation of parasites from a single Brazilian village [13]. The combined allele frequency of these four resistance alleles was 1.85%.

OXA was widely used in the New World, where only *S. mansoni* is present, during the 1970s to early 2000's [14], so the OXA-R alleles we observed in Brazil [13] could conceivably have resulted from drug selection. In contrast, OXA saw minimal usage in Africa where both

S. haematobium and *S. mansoni* are present. In South America, an estimated 6 million doses of OXA were used to treat the 7 million people infected with *S. mansoni* prior to 1987 [15]. In contrast, 3 million doses had been used in Africa (Egypt, Ethiopia, Ivory Coast, Kenya, Malagasy, Malawi, Rwanda-Burundi, South Africa, Sudan, Tanzania, Uganda, Zambia, Zaire) and the Middle East (Arabian penensula) [15], where an estimated ~60 million people were infected with *S. mansoni* [16]. Hence, OXA selection was minimal in Africa compared with South America. The central aim of this paper is to determine whether *de novo* mutation or standing variation best explains OXA-R in *S. mansoni* populations. To do this we examined *SmSULT-OR* genetic variation in Old World parasite populations where OXA treatment has been minimal.

Results

Samples

We sequenced exomes from 92 miracidia from Tanzania (n = 57), Niger (n = 10), and Senegal (n = 25) from the SCAN collection at the Natural History Museum [17], with each miracidium derived from a different patient. In addition, we included 112 samples collected from Oman: these included 86 single worms and 26 pools of cercariae from single infected snails. The Omani samples were derived from 11 human infections, 2 naturally infected rodents, and 26 naturally infected snails (Table 1). We obtained an average of 15.7 ± 8 (mean ± standard deviation, s.d.) million reads per library (S1 Table). A very high proportion of these reads were mapped to the reference genome (91.9 ± 6.29 (s.d.) % on average). We captured on average 99.08 ± 3.92 (s.d.) % of the bait regions. We obtained an average read depth of 64.40 ± 32.27 (s.d.) in the bait regions. For *SmSULT-OR*, we obtained sequence across 95.06 ± 6.54 (s.d.) % of the coding sequence with an average read depth of 43.09 ± 26.96 (s.d.).

We compared these Old World parasite samples with: (i) the OXA-R parasite from Puerto Rico (HR) carrying the causative p.E142del mutation identified previously [10], for which whole genome sequence is available, and (ii) published *SmSULT-OR* sequences from a single Brazilian location (n = 189) [13] (Table 1).

New and known mutations identified in the *SmSULT-OR* gene

We identified a total of 85 *SmSULT-OR* mutations across all three populations (West Africa, East Africa and Middle East), including 76 coding mutations and 9 non-coding mutations (S2

Table 1. Summary of the samples.

Location	Sample size	Parasite stage sampled	Host	Sequence	Notes[b]
South America (Brazil)[a]	189	Miracidia	Human	Sanger sequence of *SmSULT-OR* exons	1–11 miracidia from each of 49 people from single village [13]
West Africa (Niger/Senegal)	25/10	Miracidia	Human	Exome sequence	1 miracidia from each person
East Africa (Tanzania)	57	Miracidia	Human	Exome sequence	1 miracidia from each person
Middle East (Oman)	112	Miracidia/adult worms/cercariae	Human/Rat/Snail	Exome sequence	1–53 parasite genotypes analyzed per host from 11 people, 2 rats and 26 snails from 8 locations
Caribbean (Puerto Rico)[a]	1	Adult worms	Hamster	Genome sequence	OXA-resistant parasite used in a genetic cross [10]

[a] Data from Brazil from a previous study [13] added for comparison

[b] see S3 Table for detailed information on sample locations

Table). Exon 1 carried 24 mutations. These included 13 non-synonymous single nucleotide polymorphisms (SNPs), 8 synonymous SNPs, one duplication, one deletion, and one premature stop codon. Exon 2 carried 52 mutations: 26 non-synonymous SNPs, 23 synonymous and three deletions. There were no differences in the density of mutations in the two exons (Fisher's exact test, d.f. = 1, p = 0.09). The overall transition/transversion ratio was 1.94. Forty-three percent (37/85 mutations) showed an average allele frequency greater than 5%.

Among the 85 mutations, we have previously found three coding and one non-coding mutations in the New World (Caribbean and Brazil): the substitution p.P67L (g.200C>T), the deletion p.E142del (g.4348_4350delGAA), the substitution p.L256W (g.4691T>G) and the 3' untranslated region (UTR) substitution g.4720C>T [10,13]. Just one of these mutations (p.E142del) encodes a confirmed OXA-R allele. Frequencies of all mutations in each population are shown in S2 Table.

We measured nucleotide diversity (π) in both New and Old World populations. Nucleotide diversity was the highest in Oman (π = 247.5 ± 3.4 × 10^{-4}) followed by East Africa (π = 22.8 ± 2.3 × 10^{-4}), West Africa (π = 7.2 ± 1.5 × 10^{-4}) and Brazil (π = 2.1 ± 0.3 × 10^{-4}). If drug pressure drives the changes at *SmSULT-OR*, we might expect to see signals of positive selection. We examined sequence conservation by measuring proportions of non-synonymous difference per non-synonymous site relative to synonymous changes per synonymous site (K_a/K_s) between *S. mansoni* and a *S. rodhaini* outgroup. K_a/K_s ranged from 0.20–0.24 suggesting that changes at non-synonymous sites are 4–5 times less abundant than at synonymous sites, consistent with weak purifying selection. Similarly, the ratio of non-synonymous polymorphisms per non-synonymous site relative to synonymous polymorphisms per synonymous site (pN/pS) ranged from 0.11 to 1.23 and none of these values were significantly greater than unity (Table 2). These tests did not provide evidence that *SmSULT-OR* is under positive selection.

In vitro tests and *in silico* evaluation of impact of mutations

We used three approaches to identify *SmSULT-OR* mutations that are likely to result in OXA-R.

Visual assessment. Some mutations have obvious deleterious effects on the protein by introducing premature stop codons. This is the case for three mutations identified in

Table 2. Summary statistics for coding sequence variation in *SmSULT-OR*.

Location	Nucleotide diversity (π ± standard deviation)	Dn	Pn	Ds	Ps	N	S	pN/pS	Ka/Ks
South America (Brazil)[a]	2.1 ± 0.3 × 10^{-4}	17	4	23	0	598.19	163.81	N/C	0.202
West Africa (Senegal / Niger)	7.2 ± 1.5 × 10^{-4}	19	9	23	2	600.93	164.07	1.229	0.225
East Africa (Tanzania)	22.8 ± 2.3 × 10^{-4}	18	19	23	9	593.80	162.20	0.577	0.214
Middle East (Oman)	247.5 ± 3.4 × 10^{-4}	10	15	11	36	598.80	160.20	0.111	0.243
Overall	148.1 ± N/C × 10^{-4}	10	39	11	40	583.61	157.39	0.263	0.245

[a] Data from Brazil from a previous study [13] added for comparison

Dn, Ds: Number of fixed non-synonymous and synonymous differences between species (*S. mansoni* and *S. rodhaini*)

Pn, Ps: Number of polymorphic non-synonymous and synonymous differences within species (*S. mansoni*)

N: Number of non-synonymous sites
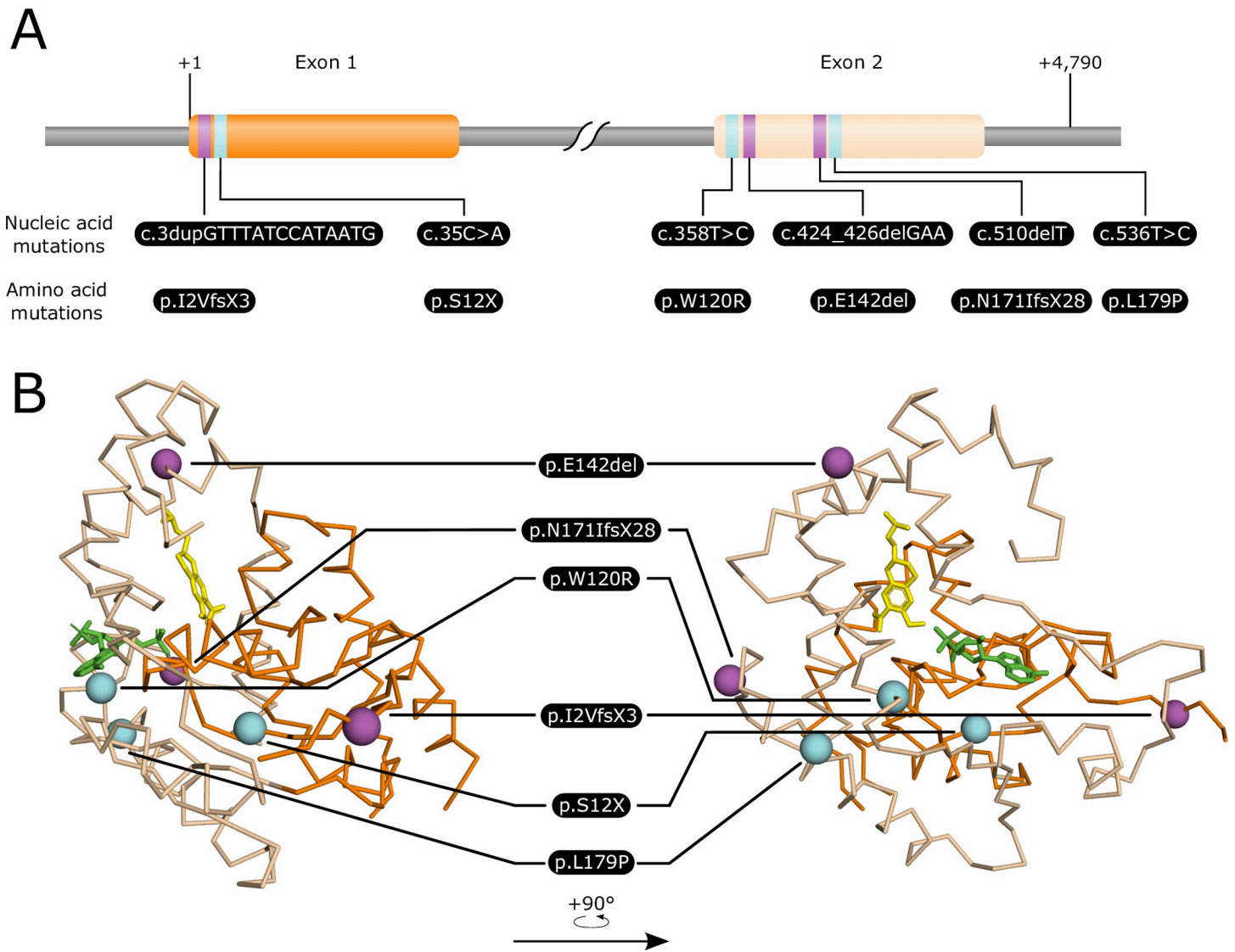
S: Number of synonymous sites

pN/pS = (Pn/N)/(Ps/S)

Ka/Ks = (Dn/N)/(Ds/S)

N/C: not computable

https://doi.org/10.1371/journal.ppat.1007881.t002

Fig 2. Mapping of the resistance mutations on the gene sequence and structure of *Schistosoma mansoni* SmSULT-OR sulfotransferase. Exon 1 and exon 2 are represented in orange and beige, respectively. Single nucleotide polymorphisms and duplication/deletion events are represented in cyan and magenta, respectively. (A) Linear representation of the *SmSULT-OR* gene showing the relative position of the mutations and their translation in amino acid sequences. (B) Positions of mutations on the SmSULT-OR protein. Oxamniquine is represented in yellow, 3'-phosphoadenosine-5'-phosphosulfate (PAPS) co-factor is represented in green.

SmSULT-OR (Fig 2). These mutations will generate truncated proteins that do not contain the active site. Other mutations, such as p.L179P, p.P225S or p.W120R, were either close enough to the OXA or the 3'-phosphoadenosine-5'-phosphosulfate (PAPS) binding sites or showed potential steric hindrance or structural deformation likely to have an effect on protein activity. Other mutations, such as p.A74T, p.P106S or p.Q176R, were on the surface of the protein, distant to any active site, and therefore unlikely to have an impact.

Functional assay of OXA binding. We used an *in vitro* OXA activation assay to functionally assay the impact of mutations on SmSULT-OR activity for four newly discovered mutations and the two known OXA-R mutations (Fig 3). The known resistant mutations, p.C35R
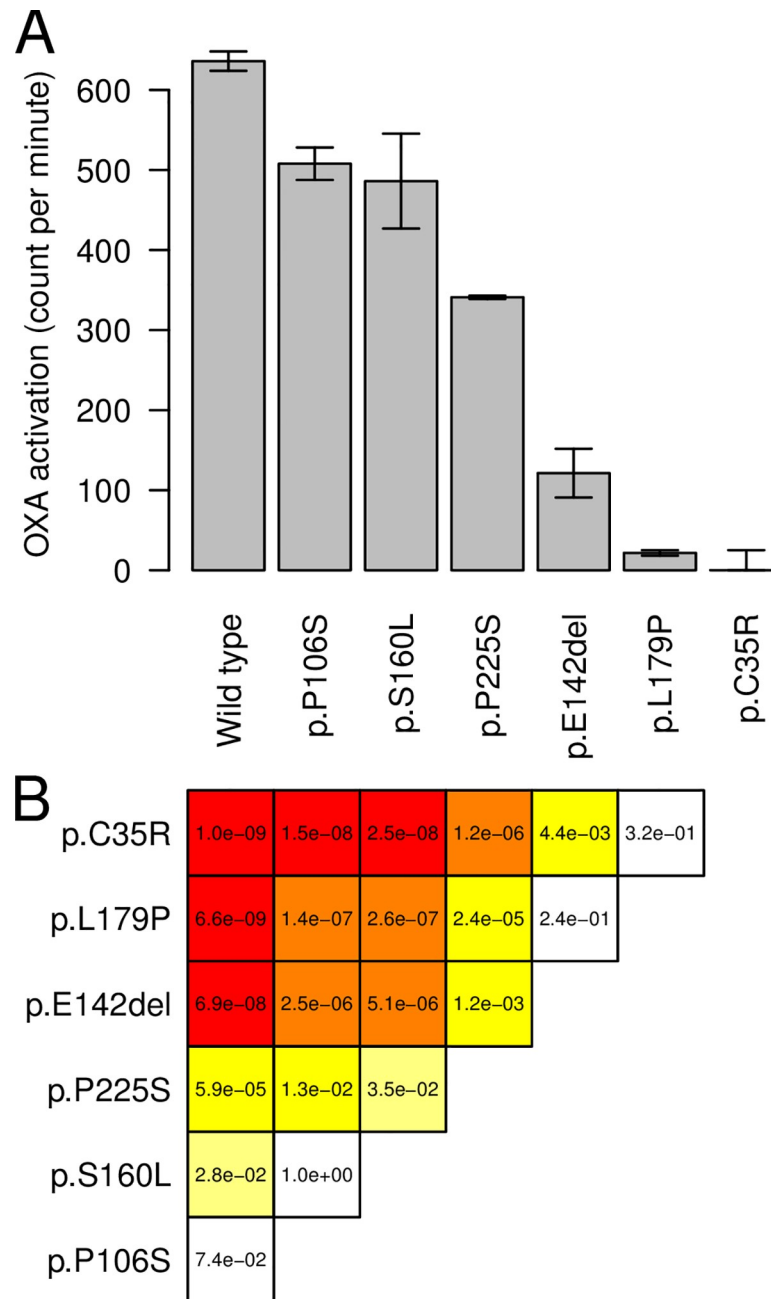
Fig 3. Enzymatic activity of recombinant _Schistosoma mansoni_ SmSULT-OR sulfotransferase expressed from different allelic variants. This _in vitro_ oxamniquine activation assay quantifies DNA-oxamniquine complexes by scintillation (counts per minute). (A) Bars show the mean of three replicates, while error bars are S.E.M. (B) The triangular matrix shows the p-values from the pairwise comparisons (Tukey's HSD). The color is proportional to the level of significance, from white (not significant) to red (most significant). Enzyme carrying known loss-of-function mutations, such as p.C35R or p.E142del, as well as a newly identified variant (p.L179P) showed no or low oxamniquine activation, while two newly identified variants (p.S160L and p.P225S) showed intermediate activation. The newly identified p.P106S did not impair oxamniquine activation.

and p.E142del, showed from zero to very low level of OXA activation, as expected [10]. p. L179P showed a similarly low level of activation. The mutations p.S160L and p.P225S showed intermediate OXA activation: we conservatively classed these as OXA sensitive alleles. p.P106S

showed similar activity to the wild type and p.S160L. We were able to produce recombinant protein p.W120R but all our attempts to fold this protein were unsuccessful. This mutation clearly has a dramatic impact on protein stability as predicted by visual inspection and thermo-dynamic modelling results (below). We considered this mutation to be OXA-R.

Thermodynamic modelling. We used thermodynamic modelling to evaluate the potential impact of substitutions on protein stability (indels such as p.E142del cannot be modelled). The difference in free enthalpy ($\Delta\Delta G$) ranged from -2.942 to 19.826 for the 37 mutations (Fig 4). Among those, p.L179P and p.C35R which have the greatest impact on OXA activation (Fig 3) showed the second and third highest $\Delta\Delta G$, respectively. p.W120R, for which we were unable to fold recombinant protein, showed the highest $\Delta\Delta G$, consistent with a dramatic impact on stability. p.S160L and p.P225S showed intermediate $\Delta\Delta G$ consistent with the results of the OXA activation assay.

In total, we identified 6 independent OXA-R mutations in Old World parasite populations examined using these three approaches (Table 3, Fig 2). These included 1 duplication, 1 known indel, 2 new indels and 3 SNPs.

Frequency of OXA-R alleles

Having determined which of the mutations identified are likely to cause OXA-R, we measured the frequency of resistance variants and the frequency of resistant parasites in each population (Table 3, Fig 5). To accurately identify resistance alleles and parasites, we phased our variant calling data on the first 3 Mb of the chromosome 6 (corresponding to 30,812 variable sites). West African samples carried only two resistance variants (p.S12X and p.E142del). East African samples carried five resistance variants, two of them at a frequency over 0.05 (p.W120R and p. N171IfsX28). The p.W120R mutation was also found at high frequency in Middle East (0.065). We found the highest frequency of OXA-R alleles in East Africa (14.91%, 17/114), followed by the Middle East (6.25%, 14/224), West Africa (4.29%, 3/70), and Brazil (1.85%, 7/378).

Most parasites carried OXA-R alleles in the heterozygous state and are predicted to be OXA sensitive. Heterozygous OXA-sensitive parasites were found in West and East Africa (8.6% and 19.3%, respectively), Middle East (12.5%) and in Brazil (1.6%). Parasites that are homozygous for OXA-R alleles at *SmSULT-OR* (but not necessarily carrying the same resistance variant) are OXA-R. We found homozygous parasites predicted to be phenotypically OXA-R at a frequency of 5.26% (3/57) in East Africa and 1.06% (2/189) in South America only (Fig 5).

Haplotype analysis of *SmSULT-OR* and flanking regions

We used our phased data to investigate the haplotypes surrounding our resistance variant (p. E142del) in the Old and New World (Caribbean) (Table 1) and to investigate whether these resistance alleles derived from a common ancestor. This is of interest because it would suggest that these alleles predate the slave trade. We identified an identical 102.5 kb haplotype block around p.E142del shared between the Caribbean sample and one of the West African samples from Niger (Fig 6A). This haplotype block contained 399 SNPs that varied in at least one of the 205 parasites examined (all samples except those from Brazil where haplotype data were not available). We used a minimum spanning network to investigate haplotype relationships using the 399 bi-allelic variants found in this 102.5 kb block across all phased haplotypes (Fig 6B). The haplotypes associated with the p.E142del OXA-R alleles found in both Puerto Rico and Niger clustered together, consistent with a common origin for West African and Caribbean p. E142del OXA-R mutations.

The p.E142del mutation was also identified in Brazil [13] but we sequenced the *SmSUL-T-OR* exons only, rather than complete exomes, for these Brazilian samples preventing
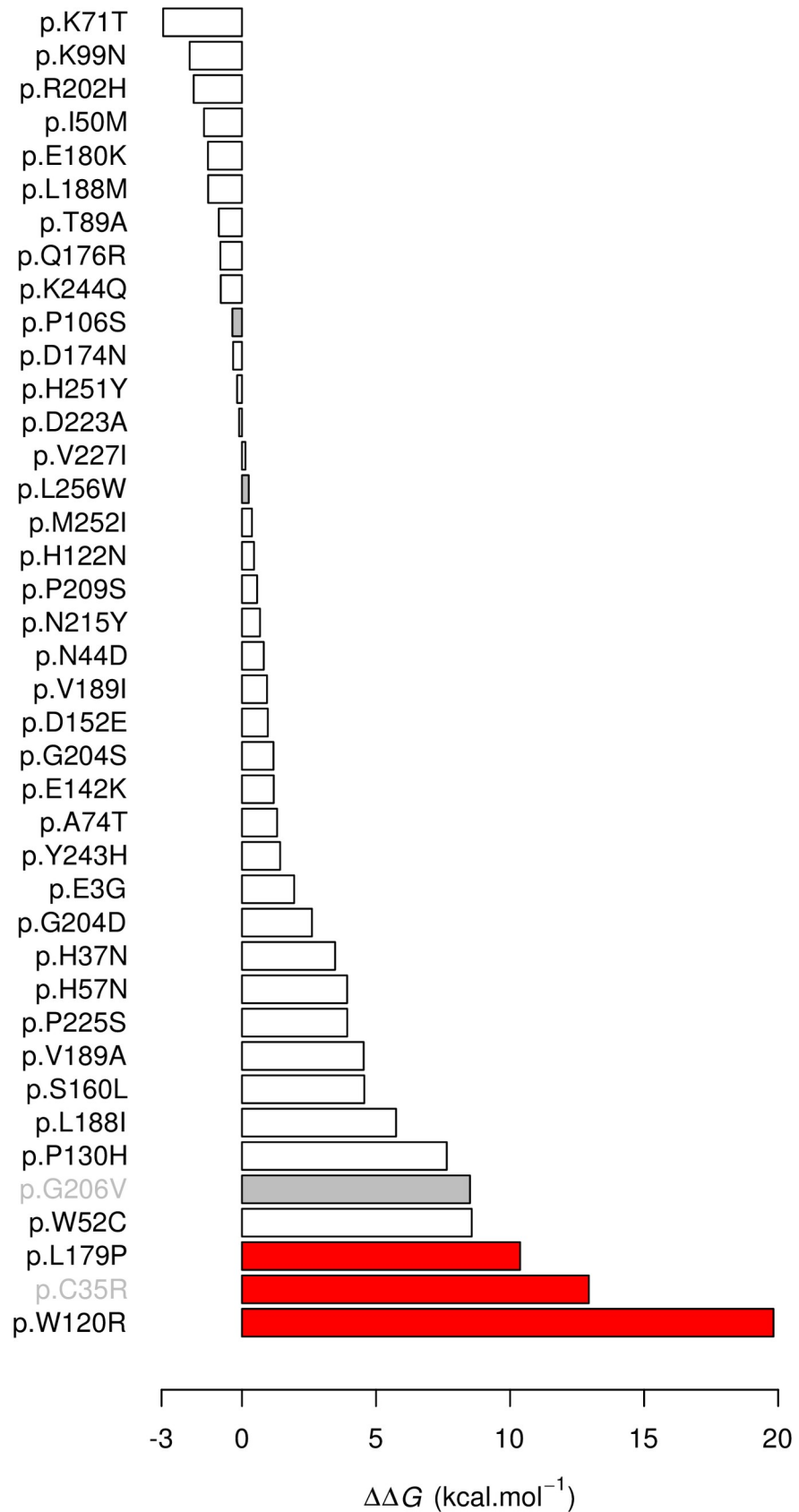
Fig 4. *In silico* evaluation of mutations on protein stability. We computed the difference in free enthalpy (*ΔΔG*) between the mutated and the wild type proteins. The higher the *ΔΔG*, the more unstable is the mutated protein. Only single amino acid changes within the resolved crystal structure were examined. For completeness, we included mutations from our current dataset and from previous studies [10,13]. Grey bars correspond to known sensitive alleles. Red bars correspond to validated resistance alleles. Grey labels correspond to mutations identified previously from South America [13].

https://doi.org/10.1371/journal.ppat.1007881.g004

reconstruction of extended haplotypes. This Brazilian p.E142del was also found associated with two mutations (p.P67L and g.4720C>T) found in West Africa but without the p.L256W found in samples from the Caribbean and Niger. Similarly, the p.E142del was also found in a Tanzanian sample but this sample had a distinct haplotype from those found in West Africa (Fig 6). These results suggest either that there are several independent origins of East and West African p.E142del mutations, or alternatively, that the p.E142del is extremely old and has distinctive flanking haplotypes resulting from recombination and mutation.

Discussion

Evidence for standing genetic variation for OXA resistance

Two lines of evidence support the view that standing variation is the source of OXA-R alleles in schistosomes. First, we have indirect evidence: OXA-R alleles are geographically widespread in Africa and the Middle East despite limited treatment with OXA (or hycanthone, its structural analog; [18]) in these regions. Furthermore, that OXA-R alleles are found at high frequency in East Africa, where there has been minimal use of OXA, while OXA-R alleles are found at low frequency in Brazil, where OXA has been used extensively, is also consistent with the idea that mutation and drift, rather than selection, explain the patterns of variation observed. The second line of evidence is more direct: we showed that an OXA-R mutation, p.E142del, with an identical 102.5 kb flanking haplotype, is sampled from both West African and Caribbean schistosomes. This result strongly suggests that p.E142del alleles were present in West Africa prior to the transatlantic slave trade (1501–1867) [19], and were transferred to

Table 3. OXA-R mutations scored in the *Schistosoma mansoni SmSULT-OR* gene and their frequency in the New and Old World.

Nucleic acid mutation	Amino acid mutation	Mutation frequency			
		South America (Brazil)[a] n = 189	West Africa (Senegal / Niger) n = 25 / n = 10	East Africa (Tanzania) n = 57	Middle East (Oman) n = 112
c.3dupGTTTATCCATAATG	p.I2VfsX3	-	-	0.0096	-
c.35C>A	p.S12X	-	0.014	-	-
c.103T>C	p.C35R	0.0027	-	-	-
c.358T>C	p.W120R	-	-	0.0510	0.065
c.424_426delGAA	p.E142del	0.008	0.035	0.0098	-
c.510delT	p.N171IfsX28	-	-	0.0625	-
c.536T>C	p.L179P	-	-	0.0089	-
Total frequency		0.0107	0.044	0.1418	0.065

The code used for nucleic acid mutations indicates the sequence type (c = coding), the position, and the mutation type (X>Y = substitution of X by Y, insN = insertion of N, delN = deletion of N, dupN = duplication of N). The code used for amino acid mutations indicates the sequence type (p = protein), the reference amino acid, the position, and finally the alternative amino acid, and when frame shift (fs) occurs, the position of the stop codon (X) after the mutation. For details about the nomenclature, see Ogino *et al.* [79].

[a] Data from Brazil were obtained in a previous study [13] added for comparison.

https://doi.org/10.1371/journal.ppat.1007881.t003
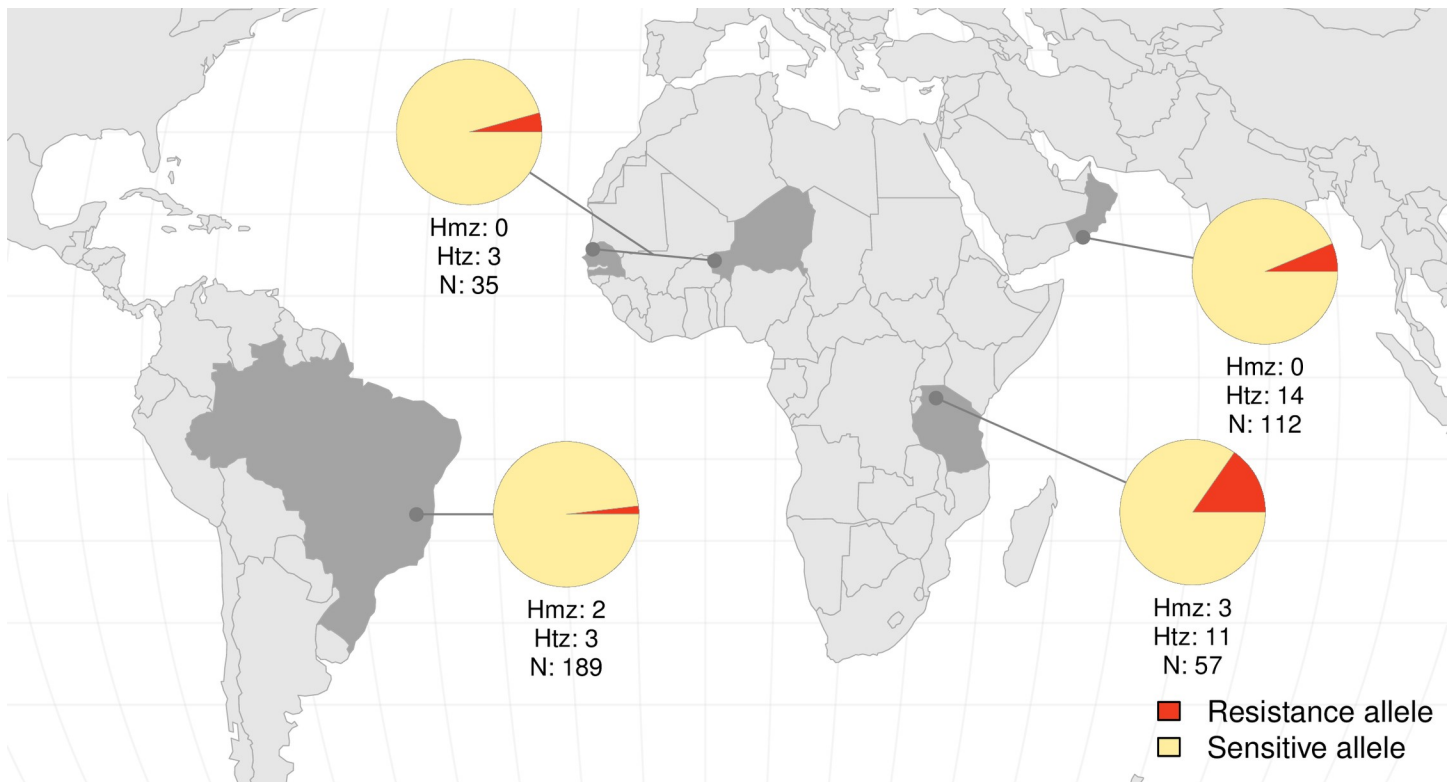
Fig 5. World map showing proportion of resistance and sensitive alleles in South America (Brazil), West Africa (Senegal and Niger), East Africa (Tanzania) and Middle East (Oman). Number of homozygous (hmz) and heterozygous (htz) parasites for resistance alleles, and total number of parasites sampled (N) are shown below the pie charts. East Africa showed the highest frequency of resistance alleles and resistant parasites. Data from Brazil from a previous study [13] was added for comparison.

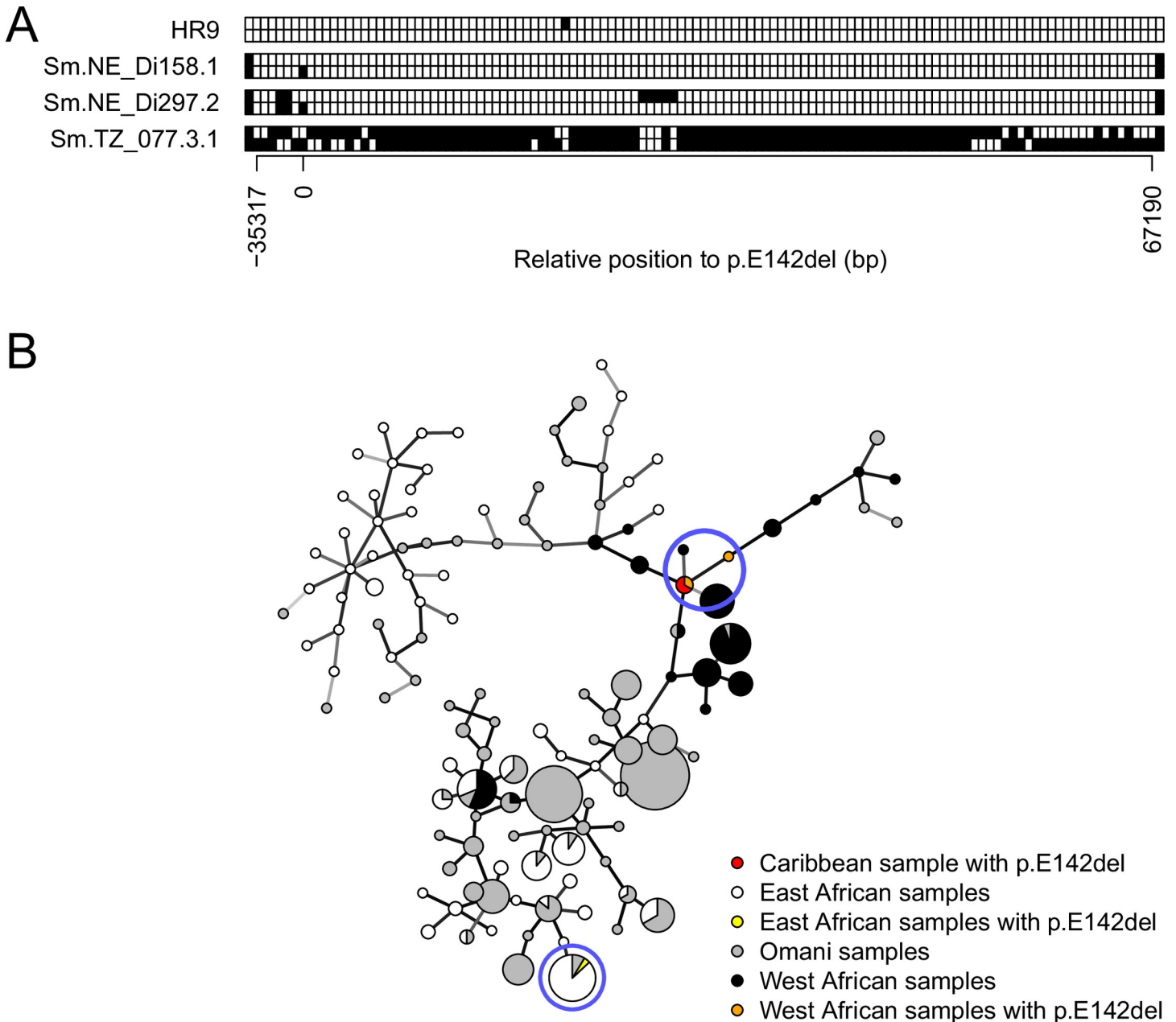https://doi.org/10.1371/journal.ppat.1007881.g005

the New World on ships carrying West African slaves. Hence, OXA-R alleles were segregating within *S. mansoni* populations at least 470 years before deployment of OXA in the 1970s [20].

Our molecular data are backed up by clinical observations from the early use of OXA. Resistant parasites were detected in Brazil in the 1970s [21,22], before any mass drug administration [14]. Similar observations of *S. mansoni* infected patients resistant to OXA treatment were made in East Africa [23]. The existence of OXA-R parasites was confirmed experimentally by infecting mice with parasites isolated from patients that were parasite positive following drug treatment [21–23]. Together these observations support the existence of segregating OXA-R alleles in schistosome populations before OXA deployment for parasite control.

Differences in OXA treatment efficacy between East and West Africa

Given the variation in OXA-R frequencies that we observe in different geographical samples, it is interesting to examine the literature on clinical treatment efficacy of OXA for treating schistosomiasis patients. Early trials of OXA to treat patients with intestinal schistosomiasis resulted in multiple treatment failures in Egypt, East Africa and South Africa [23–27]. As a consequence, the WHO recommended use of higher doses of OXA in East Africa, compared with West Africa [28]. Human host metabolism did not explain this lack of efficacy because the OXA availability in blood was the same or higher in the East African patient populations than in West African or South American populations [29,30]. Therefore, the presence of resistant parasites was the most likely explanation. OXA-R parasites were identified in Kenya by treating mice infected with parasites from patients showing poor treatment response [23]. We

A

B

- ● Caribbean sample with p.E142del
- ○ East African samples
- ● East African samples with p.E142del
- ● Omani samples
- ● West African samples
- ● West African samples with p.E142del

Fig 6. Common origin of the p.E142del mutation in the Old and New World. (A) Haplotype variation in all the samples bearing p.E142del from Caribbean (HR9), Niger (NE) and Tanzania (TZ) across a 102.5 kb region of chr. 6. Each row represents a chromosome, HR9 was used as reference, blank squares reflect HR9 allele state, and black squares correspond to the alternative allele. Relative bp position to p.E142del (0 bp) is shown on the x-axis. The first and last variants showed on the block correspond to the break of the haplotype block. The Caribbean sample (HR9) and a Nigerien sample (Sm.NE_Di158.1) share an identical haplotype block of 102.5 kb. (B) Minimum spanning network of 410 haplotypes of the 102.5 kb region previously identified. The network was built using 399 bi-allelic variants. Node size is proportional to sample size (smallest node: n = 1; biggest node n = 50). Nodes with samples carrying p.E142del are circled in blue. Caribbean and West African haplotypes carrying p.E142del clustered together indicating a common origin. The p.E142del from East Africa has different flanking haplotypes.

observed a 14.91% frequency of OXA-R alleles (and 5.26% frequency of homozygous OXA-R parasites) in East Africa compared with 4.29% frequency of OXA-R alleles (and no homozygous OXA-R parasites) in West Africa. Our results provide a molecular explanation for the poor treatment response observed in East Africa, relative to other locations.

Implications for development of OXA derivatives

OXA kills *S. mansoni* but not *S. haematobium* or *S. japonicum*. OXA derivatives are currently under development to obtain more potent molecules acting against all three schistosome species infecting humans. In fact, OXA derivatives that kill adult worms of the three species *in vitro* have now been developed [31]. The low frequency of homozygous OXA-R parasites in *S. mansoni* suggests that OXA derivatives will result in imperfect cure rate for this species, particularly in East Africa. However, the primary purpose of generating OXA derivatives is to treat *S. haematobium* and *S. japonicum*, perhaps as a partner drug for praziquantel (PZQ). Examining natural variation in homologous sulfotransferase genes from these two species (*ShSULT-OR* and *SjSULT-OR*) to better understand the potential for resistance evolution will be critical if OXA derivatives are to be developed to control these parasites.

## Implications of standing variation for drug resistance evolution

Standing variation in drug resistance genes has a major impact on how fast resistance evolves [2]. If resistance alleles are already present in the population prior to treatment, drug resistance has the potential to spread much more rapidly because there is no waiting time for a resistance mutation to appear *de novo* (Fig 7A) [3]. An increasing number of examples support this view. The nematode *Caenorhabditis elegans* carries a large diversity of *β-tubulin* alleles allowing worms to be resistant to benzimidazoles (BZ) [32]. In the filarial worm *Onchocerca volvulus*, standing variation appears to be involved in resistance to ivermectin [6]. Standing variation is also involved in herbicide resistance in plants: resistance alleles from the weed *Alopecurus myosuroides* were identified in plant collection almost 100 years before herbicides were used [33]. Antibiotic resistance provides particularly dramatic examples of standing variation: plasmids encoding multidrug resistance were found in frozen bacteria isolated from the permafrost demonstrating that antibiotic resistance plasmids were already present thousands of years ago [34,35].

We do not want to give the impression that standing variation is the only source of variation for evolution of drug resitance. In the case of other well studied helminth parasites, such as *Haemonchus contortus* and *Teladorsagia circumcincta*, resistance alleles are consistent with evolution through recurrent *de novo* mutations or very rare standing variants. While selection of rare pre-existing BZ resistance alleles was shown in French farms [36], selection of *de novo* resistance mutations, associated with different haplotypes, was shown to be common in UK farms [37].

The fact that there is strong evidence of standing variation in OXA resistance but not in BZ resistance may reflect the differences in the nature of the mutation underlying resistance. In the case of OXA, resistance is due to loss-of-function mutations resulting in non functional SmSULT-OR. As a consequence, mutations can occur in any position that disrupts substrate binding or protein structure. Hence mutations are expected to arise frequently and will exist in natural populations. In contrast, mutations underlying BZ resistance occur in specific position within the *β-tubulin* gene. As consequence, such mutations occur much more rarely. Furthermore because *β-tubulin* is an essential gene, mutations that disrupt protein structure and result in non functional proteins are strongly selected against and cannot reach high frequency.

Standing variation for OXA resistance alleles could arise by drift, or alternatively as a result of selective forces other than OXA treatment. For example, Hahnel *et al.* [32] demonsrated BZ resistance in natural population of *C. elegans* which are not targetted by BZ drugs and speculated that natural compound with similar structure to benzimidazoles could have pre-selected for benzimidazole resistance alleles. In the case of our OXA-R mutations in *SmSULT-OR*, we do not know what selective forces might be involved (see also discussion in section *Prediction of the resistance spread*).
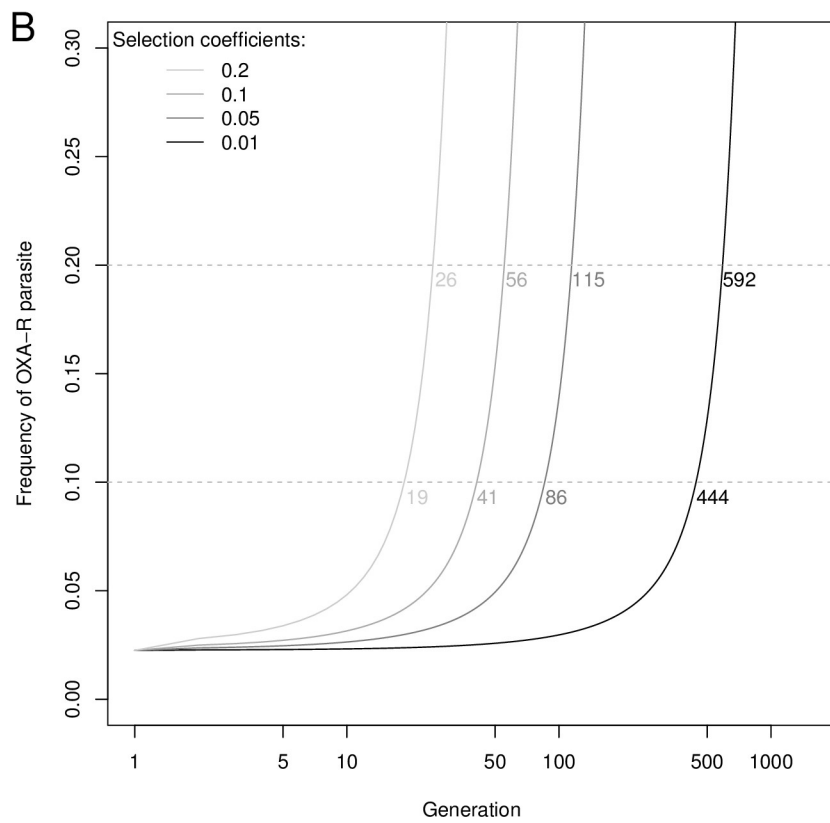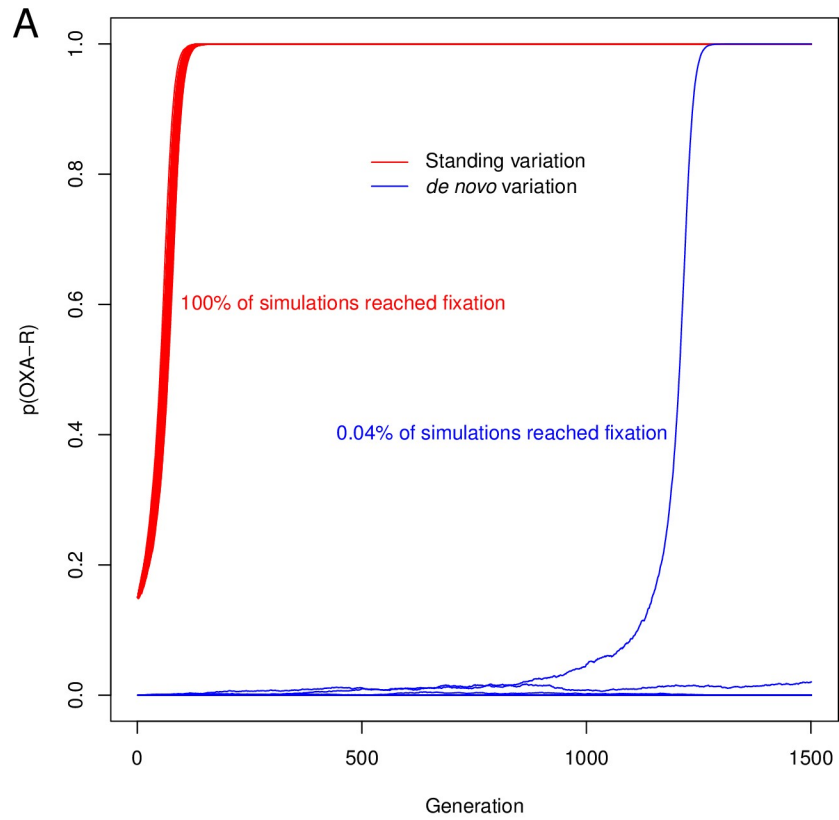
Fig 7. Impact of starting allele frequency on OXA-R allele change. (A) Monte Carlo simulation over 1,500 generations using strong selection (selection coefficient (s) = 0.1) on a population size (N) of 65,000 with a starting allele frequency (p(R)) of 0.15 for standing variation or 1/(2N) for new mutation. These simulations underestimate time to fixation for true *de novo* mutation, because we do not account for the waiting time for resistance to appear which is dependant on the rate of drug resistance mutations and N_e [3]. The starting frequency of 0.15 corresponds to the frequencies of OXA-resistance alleles observed in Kenya. (B) Change in frequency of resistant parasites (i.e. homozygotes for OXA-R alleles) from standing variation under a range of selection coefficients. The dashed lines correspond to the two thresholds (10% and 20%) at which treatment efficacy would be compromised. The numbers at the dashed lines correspond to the parasite generations needed to cross these thresholds. These predictions are deterministic, because we expect minimal stochastic variation when starting resistance allele frequencies are high.

Drug resistance in parasitic helminths infecting livestock or humans is now recognized as a potential concern for control efforts, but was neglected for a long time using the argument that the slow reproduction rate of worms would greatly delay resistance evolution [38]. The observation that resistance alleles exist as standing variation within multiple pathogen or parasite populations in the absence of drug treatment suggests previous optimism about the shelf life of drugs was poorly founded. If resistance alleles are already at relatively high frequency in populations when new drugs are deployed, resistance may evolve extremely rapidly [3,5].

Prediction of the resistance spread

We can use the frequency of resistance allele to model resistance spread in parasite populations. In the case of OXA resistance, using an N_e estimate of 65,000 [39], OXA-R alleles spread from 15% starting frequency to fixation in 170 generations in 100% of simulations when selection is strong (s = 0.1). Assuming a generation time between 3 months and 1 year, this is equivalent to a period of 42.5 to 170 years (Fig 7A). However, if OXA-R alleles were evolving *de novo*, simulations indicate that fixation will occur rarely (0.04% of the simulations) and this would take 1,197 generations (corresponding to a period of 299.25 to 1,197 years) from the time a mutation arises. We can use the observed frequency of OXA-R alleles segregating in East African populations to make predictions about the spread of phenotypically resistant parasites (i.e. those homozygous for OXA-R alleles) (Fig 7B). If we use a frequency of 10% resistant parasites to denote unacceptable efficacy (this threshold is used for antimalarial drugs [40]) then this threshold will be reached in 19 to 444 generations (4.75 to 444 years) using a range of selection coefficients (from 0.01 to 0.2). The selection coefficients driving spread of drug resistance alleles in schistosomes is not known, but the range we have used covers measured selection coefficients driving drug resistance spread in another human parasite (*Plasmodium falciparum*) [41,42]. These simulations suggest that, while OXA-R allele frequencies are relatively high in East Africa, OXA derivatives could still be extremely effective for schistosome control in the short-medium term, particularly if deployed as a combination therapy, partnered with a second drug with a different mode of action (e.g. PZQ).

Our simple models for predicting spread of OXA-R alleles under drug selection ignore fitness costs of OXA-R alleles, which might limit the rate of spread. The fitness cost of inactive SmSULT-OR in the absence of drug treatment still needs to be assessed accurately [13] but the occurrence of natural homozygotes for defective alleles suggests limited costs. The exact role of this enzyme in schistosome biology is not known. Sulfotransferases transfer the sulfo group from a donor (usually PAPS) to the substrate, resulting in inactivation or solubilization of the substrate [43]. Substrates can be endogenous signal molecules (hormones, neurotransmitters) or exogenous chemicals. In the latter case, sulfotransferases play a role in detoxifying natural or synthetic toxins [43], but sometimes they do the opposite by activating pro-drugs as observed with the anti-tumor drug N-benzyl indolecarbinols [44] or OXA [10,45]. Because the same enzyme can sulfate both endogenous and exogenous compounds, it is often difficult to

determine their exact biological role. In schistosomes, we speculate that SmSULT-OR could have roles in regulating endogenous molecules as well as detoxifying chemicals present in their environment (*i.e.* in the feces, water, snail or human blood, etc.).

Molecular markers for resistance surveillance

Molecular markers of OXA-R now allow efficient monitoring of the distribution of OXA-R alleles in schistosome populations. This approach is widely used for other parasites and pathogens such as malaria parasite [46] and HIV [47]. For schistosome, collections such as SCAN which contain thousands of parasite samples from multiple locations [17] will be extremely valuable for this work: regular sampling in different endemic regions will help to update the resistance landscape. However, not all identified *SmSULT-OR* mutations will lead to resistance, so it is critical that molecular screening is paired with functional evaluation or computational prediction, as we have done here.

PZQ is the only drug currently available to treat schistosomiasis and effort in mass drug administration has recently been expanded 10 fold [48], with a target of administering 250 million treatments per year, increasing selective pressure for PZQ resistance. We suspect that standing variation for PZQ resistance is also likely, because this trait is easy to select in the laboratory. This has been done by several groups using independent laboratory populations, and resistance has spread within very few generations [49–54], strongly suggesting that resistance alleles were already present in these laboratory populations. Furthermore, the laboratory populations used for selection were isolated before PZQ was available [23,55] and so were not exposed to drug selection. Molecular markers for monitoring the frequency of PZQ-R in natural parasite populations would be extremely valuable, as demonstrated by our work on OXA-R.

Materials and methods

Ethics statement

African samples. Samples from Senegal and Niger were collected as part of the EU-CONTRAST project [56], a multidisciplinary alliance to optimize schistosomiasis control and transmission surveillance in sub-Saharan Africa, as detailed in Webster *et al*. [57] with ethical approval granted by ethical committees of the Ministry of Health (Dakar, Senegal), and the Niger National Ethical Committee (Niamey, Niger) with additional ethical approval obtained from the St Mary's Hospital Local Ethics Research Committee, R&D office (part of the Imperial College Research Ethics Committee (ICREC; EC no. 03.36. R&D no. 03/SB/033E)), in combination with the ongoing CONTRAST and Schistosomiasis Control Initiative (SCI) activities.

For the Tanzanian samples collected as part of the Schistosomiasis Consortium for Operational Research and Evaluation (SCORE), ethical approvals were granted by the Imperial College Research Ethics Committee and the SCI Ethical approval (EC no. 03.36. R&D no. 03/SB/033E); the National Institute for Medical Research (NIMR, reference no. NIMR/HQ/R.8a/Vol. IX/1022); University of Georgia Institutional Review Boards, Athens, GA (2011-10353-1).

Following routine procedures in the field, the objectives of the study were first explained to the local village chiefs and political and religious authorities who gave their consent to conduct the study. Written consent for the schoolchildren to participate in longitudinal monitoring of the national control programme for schistosomiasis was given by head teachers, and/or village chiefs where there were no schools, due to the fact that in African villages, written consent of the child's guardian is often very difficult to obtain owing to the associated impoverished conditions and often low literacy. Each individual child also gave verbal consent before

recruitment. Following sampling, a praziquantel treatment (40 mg.kg^{-1}) was offered to infected participants.

Omani samples. We obtained ethical clearance from the Sultan Qaboos University and the Ministry of Health of Oman to use positive stool samples for schistosomiasis collected by the Ministry of Health of Oman during epidemiological screening. Ethical approval was given by the Medical Research and Ethical Committee (MREC) of the Non-Communicable Disease Control Section of the Directorate General of Health Affairs, Headquarters, Ministry of Health, Oman (no. MH/DGHA/DSDC/NCD/R&S/167/01).

We obtained ethical approvals of animal studies in France from the French Ministère de l'Éducation Nationale, de la Recherche et de la Technologie, from the French Ministère de l'Agriculture et de la Pêche (agreement no. A 66040), and from the French Direction Départementale de la Protection des Populations (no. C 66-136-01 with prefectoral order no. 2012-201-0008). Certificate for animal experimentation was given to H.M. (authorization no. C 66.11.01; articles no. R 214–87, R 214–122 and R 215–10). Housing, breeding and animal care followed the guidelines of the French CNRS. The different protocols used in this study had been approved by the French veterinary agency from the DRAAF Languedoc-Roussillon (Direction Régionale de l'Alimentation, de l'Agriculture et de la Forêt), Montpellier, France (authorization no. 007083).

Sampling

African samples. *S. mansoni* miracidia were collected from individual patients in two different West African countries in 2007: 27 patients from two locations in Senegal, and 17 patients from two locations in Niger (S3 Table) [57,58]. Additionally, *S. mansoni* miracidia were collected as part of the SCORE program from 64 children in seven villages on the shores of Lake Victoria in Tanzania (East Africa) in January 2012 (S3 Table) [59].

Collection of miracidia was performed as previously described [57,60]. Briefly, individual stool samples from positive patients were homogenized through a mesh and washed through with water. The content was then transferred in a Pitchford funnel assembly, washed with additional water, and the filtered homogenate was drained into a Petri dish. The Petri dish containing the homogenate was then left in bright ambient light (not direct sunlight) to allow hatching of miracidia. Miracidia were visualized under a dissecting microscope and individually captured in 3–5 μL of water using a micropipette. Miracidia were then pipetted individually onto Whatman FTA cards for DNA preservation. Cards were allowed to dry for 1 hour and then stored for future research [57,58,60].

Omani samples. The Omani schistosome samples were collected during 8 field trips, from 2001 to 2015, in different areas in Dhofar, Oman and originated from either *Homo sapiens*, *Rattus rattus* or *Biomphalaria pfeifferi* (S3 Table): 11 patients from 4 localities, 2 rats from 2 localities, 26 snails from 3 localities.

We obtained schistosome adult worms from naturally infected rats and from laboratory infected mice. The mice were infected with cercariae from naturally infected snails or from laboratory snails infected with miracidia isolated from stool samples. We collected miracidia from stool samples as described in Moné *et al.* [61]. We exposed *Biomphalaria pfeifferi* snails to miracidia as described in Mouahid *et al.* [62]. We maintained laboratory and naturally infected snails at constant temperature (26˚C) and balanced photoperiod (12h light / 12h dark) and fed them with fresh lettuce *ad libitum*. Prior to infection, we anaesthetized the mice by intraperitoneal injection of an anesthetic solution (0.01 mL.g^{-1} of mouse body weight). We prepared the anaesthetic solution using 0.5 mL of Rompun (20 mg.mL^{-1}; Bayer) and 1 mL of Imalgène (100 mg.mL^{-1}; Rhône Mérieux) diluted in 8.5 mL of autoclaved NaCl 8.5 ‰.

Abdomens of anesthetized mice were shaved and exposed during 1 hour to cercariae shed from laboratory or naturally infected snails. Cercariae from naturally infected snails were kept in 95˚ ethanol when the number of cercariae was too low for mouse infection. Laboratory infected mice or naturally infected rats were euthanized by intraperitoneal injection of sodium pentobarbital solution (1.1 mL diluted in 10 mL of ethanol 10%). We recovered adult worms by perfusion [63]; the worms were washed in NaCl 8.5 ‰, and single worms (female or male) were fixed in 95˚ ethanol and preserved at -20˚C. Worms and cercariae were washed in 1X TE buffer for 1 hour prior to DNA extraction.

DNA processing and library preparation

We amplified DNA and sequenced exomes from single miracidia preserved on FTA cards (1 miracidium per patient) for African samples or from DNA extracted from worms or cercariae for Omani samples. FTA-preserved samples were processed following our published protocol [64]. Briefly we punched a 2 mm disc containing the miracidium from the FTA card, washed it with the FTA Purification Reagent (GE Healthcare Life Sciences), rinsed it twice with TE^{-1} buffer, and finally dried it. DNA from single worms and cercariae were extracted using the DNeasy Blood and Tissue kit (Qiagen) following the tissue protocol with an incubation time of 2h at 56˚C and an elution in 200 μL. Samples were quantified using the Qubit dsDNA HS assay kit (Invitrogen).

We then performed whole genome amplification (WGA) on each FTA punch or on 2 μL or 4 μL of DNA solution using the Illustra GenomiPhi V2 DNA Amplification kit (GE Healthcare Life Sciences). Amplified DNA was purified with the SigmaSpin^{TM} Sequencing reaction Clean-up (Sigma-Aldrich), following the manufacturer protocol. We quantified purified samples using the Qubit dsDNA BR assay (Invitrogen). DNA samples that failed WGA were cleaned up and concentrated using the Genomic DNA Clean & Concentrator kit (Zymo Research) following the manufacturer protocol and these samples with concentrated DNA were used to perform a new WGA.

Because WGA reactions are non-specific, we performed qPCR on each sample from FTA cards to quantify proportion of schistosome DNA (i.e., check for an excess of amplified DNA from the environment (water, fecal matter, etc.)). We excluded samples that showed less than 100 schistosome genomes copies in 20 ng of DNA which were not suitable for subsequent exome capture. We assessed genome copies by quantifyng the *S. mansoni α-tubulin* single copy gene [64].

We prepared exome capture libraries on the selected African samples and on all the successfully amplified Omani samples using the pre-capture pooling method of the SureSelect XT^{2} Target Enrichment System (Agilent) [64]. Finally, we sequenced libraries on a HiSeq 2500 and data were demultiplexed using the Casava pipeline. Raw sequencing data are accessible from the NCBI Sequence Read Archive under BioProject accession numbers PRJNA439266, PRJNA560069, PRJNA560070.

Alignment, variant calling and phasing

We aligned sequences as described in Le Clec'h *et al.* [64]. Briefly, we aligned data against the v5 *S. mansoni* genome using BWA and SAMtools, realigned around indels using GATK, PCR duplicates were marked using picard and Q-score were recalibrated using GATK.

To identify variants in the *SmSULT-OR* (Smp_089320) gene, we performed a variant calling using FreeBayes (v1.1.0-46-g8d2b3a0) [65] in the first 3 Mb of the chromosome 6. We used a minimum base quality (-q) of 20 and a minimum mapping quality (-m) of 30. We included mutation sites identified previously [10,13] in order to genotype these sites specifically for

further comparison. We defined 4 populations: Caribbean (HR), West Africa (SN and NE), East Africa (TZ) and Middle-East (OM). We excluded variants supported by less than 4 reads using VCFtools (v0.1.14) [66]. We then used vcf-subset from VCFtools to remove sites with reference allele only. We adjusted variant positions using vt normalize (v0.5772-60f436c3) [67]. We finally used Beagle (v4.1) [68] to first estimate genotypes and to then phase data. This approach resolves haplotypes using statistical models rather than by direct linkage of SNPs on short reads. Unphased variant calling data and scripts used to analyze the data are available on Zenodo (DOI: 10.5281/zenodo.2850876 and 10.5281/zenodo.3370039, respectively).

Genetics analysis

We performed all the genetics analysis using phased exome data. We handled the VCF data in R (R v3.5.0 [69]) using the vcfR package (v1.8.0.9) [70]. We determined the longest haplotype block from the samples carrying the p.E142del using a custom R code after filtering invariable sites in these samples. We built the haplotype network using the R package poppr (v2.8.1) [71,72] after selecting variants within the longest haplotype coordinates showing no more than 20% of missing data. We computed the frequency of the OXA-R allele using a custom R script and functional annotation obtained through a custom bash script.

We generated haplotype sequences of the coding sequence only for the Old World samples using the updated phased VCF data exported from R and the BCFtools (v1.2) [73]. We generated haplotypes of the South American samples from previous genotyping using a custom bash script. We included the *SmSULT-OR* homologue sequence from *S. rodhaini* previously identified [13]. We aligned the haplotypes using Clustal Omega (v1.2.2) [74]. We determined nucleotide diversity, number of synonymous and non-synonymous sites using DnaSP software (v6.12.01) [75]. Populations were defined as described in the previous section. Scripts used to analyze the data are available on Zenodo (DOI: 10.5281/zenodo.3370039).

Recombinant SmSULT-OR protein production and OXA activation assay

We produced recombinant SmSULT-OR proteins following Chevalier *et al.* [13]. Briefly, mutations were introduced in the cloned *SmSULT-OR* gene sequence (Smp_089320; GenBank accession no. HE601629.1) and introduced into *Escherichia coli* for protein production. Proteins were extracted from bacterial culture and purified on an affinity chromatography column. His-tag was removed using Tobacco etch virus (TEV) protease, the solution was dialyzed overnight and passed through affinity chromatography again to remove His-tag and TEV protease. The sample was loaded onto a GE-pre-packed Q anion exchange column and eluted, and the pooled fractions were dialyzed overnight. The protein was finally concentrated at 10 mg.mL^{-1}.

We tested the impact of a subset of the mutations observed on OXA activation using an *in vitro* assay [13]. We tested the ability of recombinant SmSULT-OR proteins to activate OXA in a protease inhibitor cocktail with sheared *S. mansoni* gDNA as a final target for the tritiated OXA pre-mixed with ATP, MgCl$_2$ and PAPS co-factor. After 2.5 h of incubation at 37˚C, the reaction was stopped and DNA was extracted three times. Radioactivity in the remaining aqueous phase containing the DNA and in water (blank) was counted in a liquid scintillation spectrometer. Blank values were subtracted from sample values. We performed three independent reactions for each recombinant protein.

Evaluation of impact of mutations on protein stability

We first visually assessed the scored mutations on the SmSULT-OR protein structure (PDB code 4MUB [10]) using the mutagenesis function of PyMol software (v2.3.0; Schrödinger, LLC).

We then used the Rosetta package (v3.9) to test the impact of scored mutations on protein stability. We used the ddg_monomer application to compute the difference in free enthalpy ($\Delta\Delta G$) between the mutated (amino acid substitutions only) and the wild-type protein. The higher the $\Delta\Delta G$, the more unstable is the mutated protein. For this, we prepared ligand files (OAQ and A3P) by downloading SDF files containing all ligand structures from SmSULT-OR crystals available from the Protein Data Bank website. We modified SDF files to add hydrogens using Openbabel (v2.4.0) [76], and generated params files using molfile_to_params.py from Rosetta. We used the structure of SmSULT-OR to generate a constraint file using the minimize_with_cst application and the convert_to_cst_file.sh scripts. We removed ligand information from the constraint file. For each mutation, we finally generated a resfile and ran the ddg_monomer application with 50 iterations and options adapted from the protocol 16 of Kellogg *et al.* [77]. Scripts used to run this analysis are available on Zenodo (DOI: 10.5281/zenodo.3370039).

Statistical analysis

We conducted statistical analyses using R (v3.5.0) [69]. We performed simulations using a modified version of the driftR simulator [78]. The OXA activation assay data were tested for normality using a Shapiro-Wilk and compared by a parametric ANOVA followed by a Tukey post-hoc test.

Supporting information

S1 Table. Library sequencing statistics.
(XLSX)

S2 Table. Mutations scored in samples from West Africa (Senegal and Niger), East Africa (Tanzania) and Middle East (Oman).
(XLSX)

S3 Table. Sampling information for the schistosome material used.
(XLSX)

Acknowledgments

Author Contributions

References

1. Barrett RDH, Schluter D. Adaptation from standing genetic variation. Trends in ecology & evolution. 2008; 23: 38–44. https://doi.org/10.1016/j.tree.2007.09.008 PMID: 18006185

2. Hermisson J, Pennings PS. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. Genetics. 2005; 169: 2335–2352. https://doi.org/10.1534/genetics.104.036947 PMID: 15716498

3. Messer PW, Petrov DA. Population genomics of rapid adaptation by soft selective sweeps. Trends Ecol Evol. 2013; 28: 659–669. https://doi.org/10.1016/j.tree.2013.08.003 PMID: 24075201

4. Orr HA, Betancourt AJ. Haldane's sieve and adaptation from the standing genetic variation. Genetics. 2001; 157: 875–884. PMID: 11157004

5. Hawkins NJ, Bass C, Dixon A, Neve P. The evolutionary origins of pesticide resistance. Biological reviews of the Cambridge Philosophical Society. 2018; https://doi.org/10.1111/brv.12440 PMID: 29971903

6. Doyle SR, Bourguinat C, Nana-Djeunga HC, Kengne-Ouafo JA, Pion SDS, Bopda J, et al. Genome-wide analysis of ivermectin response by *Onchocerca volvulus* reveals that genetic drift and soft selective sweeps contribute to loss of drug sensitivity. PLoS neglected tropical diseases. 2017; 11: e0005816. https://doi.org/10.1371/journal.pntd.0005816 PMID: 28746337

7. Diawara A, Schwenkenbecher JM, Kaplan RM, Prichard RK. Molecular and biological diagnostic tests for monitoring benzimidazole resistance in human soil-transmitted helminths. The American journal of tropical medicine and hygiene. 2013; 88: 1052–1061. https://doi.org/10.4269/ajtmh.12-0484 PMID: 23458960

8. Pullan RL, Smith JL, Jasrasaria R, Brooker SJ. Global numbers of infection and disease burden of soil transmitted helminth infections in 2010. Parasites & vectors. 2014; 7: 37. https://doi.org/10.1186/1756-3305-7-37 PMID: 24447578

9. Pica-Mattoccia L, Novi A, Cioli D. Enzymatic basis for the lack of oxamniquine activity in *Schistosoma haematobium* infections. Parasitol Res. 1997; 83: 687–689. https://doi.org/10.1007/s004360050320 PMID: 9272559

10. Valentim CLL, Cioli D, Chevalier FD, Cao X, Taylor AB, Holloway SP, et al. Genetic and molecular basis of drug resistance and species-specific drug action in Schistosome parasites. Science. 2013; 342: 1385–1389. https://doi.org/10.1126/science.1243106 PMID: 24263136

11. Rogers SH, Bueding E. Hycanthone resistance: development in *Schistosoma mansoni*. Science. 1971; 172: 1057–1058. https://doi.org/10.1126/science.172.3987.1057 PMID: 5103321

12. Pica-Mattoccia L, Dias LC, Cioli D. Genetic complementation analysis of two independently isolated hycanthone-resistant strains of *Schistosoma mansoni*. Mem Inst Oswaldo Cruz. 1992; 87 Suppl 4: 211–214. https://doi.org/10.1590/S0074-02761992000800032 PMID: 1343897

13. Chevalier FD, Le Clec'h W, Eng N, Rugel AR, Assis RR de, Oliveira G, et al. Independent origins of loss-of-function mutations conferring oxamniquine resistance in a Brazilian schistosome population. Int J Parasitol. 2016; 46: 417–424. https://doi.org/10.1016/j.ijpara.2016.03.006 PMID: 27073078

14. Coura JR, Amaral RS. Epidemiological and control aspects of schistosomiasis in Brazilian endemic areas. Mem Inst Oswaldo Cruz. 2004; 99: 13–19. https://doi.org/10.1590/S0074-02762004000900003 PMID: 15486629

15. Foster R. A review of clinical experience with oxamniquine. Transactions of the Royal Society of Tropical Medicine and Hygiene. 1987; 81: 55–59. https://doi.org/10.1016/0035-9203(87)90282-3 PMID: 3127959

16. Hotez PJ, Bundy DA, Beegle K, Brooker S, Drake L, de Silva N, et al. Disease Control Priorities in Developing Countries. In: Jamison D, Breman J, AR, M, et al., editors. 2nd ed. Washington (DC): Oxford University Press & World Bank; 2006. pp. 467–482. Available: http://www.ncbi.nlm.nih.gov/books/NBK11748/

17. Emery AM, Allan FE, Rabone ME, Rollinson D. Schistosomiasis collection at NHM (SCAN). Parasit Vectors. 2012; 5: 185. https://doi.org/10.1186/1756-3305-5-185 PMID: 22943137

18. Cioli D, Pica-Mattoccia L, Moroni R. *Schistosoma mansoni*: hycanthone/oxamniquine resistance is controlled by a single autosomal recessive gene. Exp Parasitol. 1992; 75: 425–432. https://doi.org/10.1016/0014-4894(92)90255-9 PMID: 1493874

19. Eltis D, Richardson D, Blight DW. Atlas of the Transatlantic Slave Trade [Internet]. Yale University Press; 2010. Available: http://www.jstor.org/stable/j.ctt5vm1s4

20. Katz N, Pellegrino J, Grinbaum E, Chaves A, Zicker F. Preliminary clinical trials with oxamniquine, a new antischistosomal agent. Rev Inst Med Trop Sao Paulo. 1973; 15: 25–29. PMID: 4574207

21. Katz N, Dias EP, Araújo N, Souza CP. Estudo de uma cepa humana de *Schistosoma mansoni* resistente a agentes esquistossomicidas. Rev Soc Bras Med Trop. 1973; 7: 381–387. https://doi.org/10.1590/S0037-86821973000600008

22. Dias L. C. de S. Pedro R. J. Rigo E. Goto M. M. F. Mafra G. L. Linhagem humana de *Schistosoma mansoni* resistente a esquistossomicidas. Revista de Saúde Pública. 1978; 12: 110–110. https://doi.org/10.1590/S0034-89101978000100013 PMID: 675134

23. Coles GC, Mutahi WT, Kinoti GK, Bruce JI, Katz N. Tolerance of Kenyan *Schistosoma mansoni* to oxamniquine. Transactions of the Royal Society of Tropical Medicine and Hygiene. 1987; 81: 782–785. https://doi.org/10.1016/0035-9203(87)90032-0 PMID: 3130689

24. Gupta KK. Schistosoma mansoni treatment with oral oxamniquine in Zambia. East African medical journal. 1984; 61: 641–644. PMID: 6399028

25. Omer AH. Oxamniquine for treating *Schistosoma mansoni* infection in Sudan. British medical journal. 1978; 2: 163–165. https://doi.org/10.1136/bmj.2.6131.163 PMID: 678831

26. Ongom VL, Wamboka GW, Kadil AUK. Oxamniquine (UK 4271): a potential antischistosomal drug in the treatment of *Schistosoma mansoni* infections in Uganda. The East African medical journal. 1976; 53: 505.

27. Pitchford RJ, Lewis M. Oxamniquine in the treatment of various schistosome infections in South Africa. S Afr Med J. 1978; 53: 677–680. PMID: 354047

28. World Health Organization. WHO model prescribing information: drugs used in parasitic diseases [Internet]. World Health Organization; 1995. Available: http://apps.who.int/medicinedocs/en/d/Jh2922e/

29. Daneshmend TK, Homeida MA. Oxamniquine pharmacokinetics in hepatosplenic schistosomiasis in the Sudan. The Journal of antimicrobial chemotherapy. 1987; 19: 87–93. https://doi.org/10.1093/jac/19.1.87 PMID: 3104279

30. Kokwaro GO, Taylor G. Oxamniquine pharmacokinetics in healthy Kenyan African volunteers. East African medical journal. 1991; 68: 359–364. PMID: 1935730

31. Rugel A, Tarpley RS, Lopez A, Menard T, Guzman MA, Taylor AB, et al. Design, Synthesis, and Characterization of Novel Small Molecules as Broad Range Antischistosomal Agents. ACS medicinal chemistry letters. 2018; 9: 967–973. https://doi.org/10.1021/acsmedchemlett.8b00257 PMID: 30344901

32. Hahnel SR, Zdraljevic S, Rodriguez BC, Zhao Y, McGrath PT, Andersen EC. Extreme allelic heterogeneity at a *Caenorhabditis elegans* beta-tubulin locus explains natural resistance to benzimidazoles. PLoS pathogens. 2018; 14: e1007226. https://doi.org/10.1371/journal.ppat.1007226 PMID: 30372484

33. Délye C, Deulvot C, Chauvel B. DNA analysis of herbarium Specimens of the grass weed *Alopecurus myosuroides* reveals herbicide resistance pre-dated herbicides. PloS one. 2013; 8: e75117. https://doi.org/10.1371/journal.pone.0075117 PMID: 24146749

34. D'Costa VM, King CE, Kalan L, Morar M, Sung WWL, Schwarz C, et al. Antibiotic resistance is ancient. Nature. 2011; 477: 457–461. https://doi.org/10.1038/nature10388 PMID: 21881561

35. Perry J, Waglechner N, Wright G. The Prehistory of Antibiotic Resistance. Cold Spring Harbor perspectives in medicine. 2016; 6. https://doi.org/10.1101/cshperspect.a025197 PMID: 27252395

36. Silvestre A, Humbert JF. Diversity of benzimidazole-resistance alleles in populations of small ruminant parasites. Int J Parasitol. 2002; 32: 921–928. https://doi.org/10.1016/s0020-7519(02)00032-2 PMID: 12062563

37. Redman E, Whitelaw F, Tait A, Burgess C, Bartley Y, Skuce PJ, et al. The emergence of resistance to the benzimidazole anthelmintics in parasitic nematodes of livestock is characterised by multiple independent hard and soft selective sweeps. PLoS Negl Trop Dis. 2015; 9: e0003494. https://doi.org/10.1371/journal.pntd.0003494 PMID: 25658086

38. Geerts S, Gryseels B. Drug resistance in human helminths: current situation and lessons from livestock. Clinical microbiology reviews. 2000; 13: 207–222. https://doi.org/10.1128/cmr.13.2.207-222.2000 PMID: 10755998

39. Crellen T, Allan F, David S, Durrant C, Huckvale T, Holroyd N, et al. Whole genome resequencing of the human parasite *Schistosoma mansoni* reveals population history and effects of selection. Scientific Reports. 2016; 6: 20954. https://doi.org/10.1038/srep20954 PMID: 26879532

40. Organization WH. Artemisinin resistance and artemisinin-based combination therapy efficacy: status report [Internet]. Organization WH, editor. World Health Organization; 2018. Available: https://apps.who.int/iris/bitstream/handle/10665/274362/WHO-CDS-GMP-2018.18-eng.pdf

41. Nair S, Williams JT, Brockman A, Paiphun L, Mayxay M, Newton PN, et al. A selective sweep driven by pyrimethamine treatment in southeast asian malaria parasites. Molecular biology and evolution. 2003; 20: 1526–1536. https://doi.org/10.1093/molbev/msg162 PMID: 12832643

42. Anderson TJC, Nair S, McDew-White M, Cheeseman IH, Nkhoma S, Bilgic F, et al. Population Parameters Underlying an Ongoing Soft Sweep in Southeast Asian Malaria Parasites. Molecular biology and evolution. 2017; 34: 131–144. https://doi.org/10.1093/molbev/msw228 PMID: 28025270

43. Negishi M, Pedersen LG, Petrotchenko E, Shevtsov S, Gorokhov A, Kakuta Y, et al. Structure and function of sulfotransferases. Archives of biochemistry and biophysics. 2001; 390: 149–157. https://doi.org/10.1006/abbi.2001.2368 PMID: 11396917

44. Rothman DM, Gao X, George E, Rasmusson T, Bhatia D, Alimov I, et al. Metabolic Enzyme Sulfotransferase 1A1 Is the Trigger for N-Benzyl Indole Carbinol Tumor Growth Suppression. Chemistry & Biology. 2015; 22: 1228–1237. https://doi.org/10.1016/j.chembiol.2015.06.025 PMID: 26364931

45. Pica-Mattoccia L, Carlini D, Guidi A, Cimica V, Vigorosi F, Cioli D. The schistosome enzyme that activates oxamniquine has the characteristics of a sulfotransferase. Mem Inst Oswaldo Cruz. 2006; 101 Suppl 1: 307–312. https://doi.org/10.1590/S0074-02762006000900048 PMID: 17308787

46. Nsanzabana C, Djalle D, Guérin PJ, Ménard D, González IJ. Tools for surveillance of anti-malarial drug resistance: an assessment of the current landscape. Malaria journal. 2018; 17: 75. https://doi.org/10.1186/s12936-018-2185-9 PMID: 29422048

47. Keating P, Pharris A, Leitmeyer K, De Angelis S, Wensing A, Amato-Gauci AJ, et al. Assessment of HIV molecular surveillance capacity in the European Union, 2016. Eurosurveillance. 2017; 22. https://doi.org/10.2807/1560-7917.ES.2017.22.49.17–00269

48. Fenwick A. Praziquantel: do we need another antischistosoma treatment? Future medicinal chemistry. 2015; 7: 677–680. https://doi.org/10.4155/fmc.15.16 PMID: 25996058

49. Fallon PG, Doenhoff MJ. Drug-resistant schistosomiasis: resistance to praziquantel and oxamniquine induced in Schistosoma mansoni in mice is drug specific. Am J Trop Med Hyg. 1994; 51: 83–88. https://doi.org/10.4269/ajtmh.1994.51.83 PMID: 8059919

50. Couto FFB, Coelho PMZ, Araújo N, Kusel JR, Katz N, Jannotti-Passos LK, et al. *Schistosoma mansoni*: a method for inducing resistance to praziquantel using infected *Biomphalaria glabrata* snails. Mem Inst Oswaldo Cruz. 2011; 106: 153–157. https://doi.org/10.1590/s0074-02762011000200006 PMID: 21537673

51. Lotfy WM, Hishmat MG, El Nashar AS, Abu El Einin HM. Evaluation of a method for induction of praziquantel resistance in Schistosoma mansoni. Pharmaceutical biology. 2015; 53: 1214–1219. https://doi.org/10.3109/13880209.2014.970289 PMID: 25609146

52. Pinto-Almeida A, Mendes T, Armada A, Belo S, Carrilho E, Viveiros M, et al. The Role of Efflux Pumps in Schistosoma mansoni Praziquantel Resistant Phenotype. PloS one. 2015; 10: e0140147. https://doi.org/10.1371/journal.pone.0140147 PMID: 26445012

53. Lamberton PHL, Faust CL, Webster JP. Praziquantel decreases fecundity in *Schistosoma mansoni* adult worms that survive treatment: evidence from a laboratory life-history trade-offs selection study. Infectious diseases of poverty. 2017; 6: 110. https://doi.org/10.1186/s40249-017-0324-0 PMID: 28622767

54. Sanchez MC, Cupit PM, Bu L, Cunningham C. Transcriptomic analysis of reduced sensitivity to praziquantel in *Schistosoma mansoni*. Molecular and biochemical parasitology. 2019; 228: 6–15. https://doi.org/10.1016/j.molbiopara.2018.12.005 PMID: 30658180

55. Lewis FA, Stirewalt MA, Souza CP, Gazzinelli G. Large-scale laboratory maintenance of *Schistosoma mansoni*, with observations on three schistosome/snail host combinations. J Parasitol. 1986; 72: 813–829. PMID: 3546654

56. Utzinger J, Brattig NW, Kristensen TK. Schistosomiasis research in Africa: how the CONTRAST alliance made it happen. Acta Tropica. 2013; 128: 182–195. https://doi.org/10.1016/j.actatropica.2013.08.011 PMID: 23973364

57. Webster BL, Webster JP, Gouvras AN, Garba A, Lamine MS, Diaw OT, et al. DNA "barcoding" of *Schistosoma mansoni* across sub-Saharan Africa supports substantial within locality diversity and geographical separation of genotypes. Acta Trop. 2013; 128: 250–260. https://doi.org/10.1016/j.actatropica.2012.08.009 PMID: 22935316

58. Gower CM, Gouvras AN, Lamberton PHL, Deol A, Shrivastava J, Mutombo PN, et al. Population genetic structure of *Schistosoma mansoni* and *Schistosoma haematobium* from across six sub-Saharan African countries: Implications for epidemiology, evolution and control. Acta Trop. 2013; 128: 261–274. https://doi.org/10.1016/j.actatropica.2012.09.014 PMID: 23041540

59. Ezeamama AE, He C-L, Shen Y, Yin X-P, Binder SC, Campbell CH, et al. Gaining and sustaining schistosomiasis control: study protocol and baseline data prior to different treatment strategies in five African countries. BMC Infect Dis. 2016; 16: 229. https://doi.org/10.1186/s12879-016-1575-2 PMID: 27230666

60. Gower CM, Shrivastava J, Lamberton PHL, Rollinson D, Webster BL, Emery A, et al. Development and application of an ethically and epidemiologically advantageous assay for the multi-locus microsatellite analysis of *Schistosoma mansoni*. Parasitology. 2007; 134: 523–536. https://doi.org/10.1017/S0031182006001685 PMID: 17096873

61. Moné H, Holtfreter MC, Allienne J-F, Mintsa-Nguéma R, Ibikounlé M, Boissier J, et al. Introgressive hybridizations of *Schistosoma haematobium* by Schistosoma bovis at the origin of the first case report of schistosomiasis in Corsica (France, Europe). Parasitology research. 2015; 114: 4127–4133. https://doi.org/10.1007/s00436-015-4643-4 PMID: 26268566

62. Mouahid G, Mintsa Nguema R, Al Mashikhi KM, Al Yafae SA, Idris MA, Moné H. Host-parasite life-histories of the diurnal vs. nocturnal chronotypes of *Schistosoma mansoni*: adaptive significance. Tropical medicine & international health: TM & IH. 2019; https://doi.org/10.1111/tmi.13227 PMID: 30851235

63. Duvall RH, DeWitt WB. An improved perfusion technique for recovering adult schistosomes from laboratory animals. Am J Trop Med Hyg. 1967; 16: 483–486. https://doi.org/10.4269/ajtmh.1967.16.483 PMID: 4952149

64. Le Clec'h W, Chevalier FD, McDew-White M, Allan F, Webster BL, Gouvras AN, et al. Whole genome amplification and exome sequencing of archived schistosome miracidia. Parasitology. 2018; 1–9. https://doi.org/10.1017/S0031182018000811 PMID: 29806576

65. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. 2012; Available: http://arxiv.org/pdf/1207.3907v2

66. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011; 27: 2156–2158. https://doi.org/10.1093/bioinformatics/btr330 PMID: 21653522

67. Tan A, Abecasis GR, Kang HM. Unified representation of genetic variants. Bioinformatics (Oxford, England). 2015; 31: 2202–2204. https://doi.org/10.1093/bioinformatics/btv112 PMID: 25701572

68. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. American journal of human genetics. 2007; 81: 1084–1097. https://doi.org/10.1086/521987 PMID: 17924348

69. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria; 2015. Available: http://www.R-project.org/

70. Knaus BJ, Grünwald NJ. VCFR: a package to manipulate and visualize variant call format data in R. Molecular ecology resources. 2017; 17: 44–53. https://doi.org/10.1111/1755-0998.12549 PMID: 27401132

71. Kamvar ZN, Tabima JF, Grünwald NJ. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ. 2014; 2: e281. https://doi.org/10.7717/peerj.281 PMID: 24688859

72. Kamvar ZN, Brooks JC, Grünwald NJ. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Frontiers in genetics. 2015; 6: 208. https://doi.org/10.3389/fgene.2015.00208 PMID: 26113860

73. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011; 27: 2987–2993. https://doi.org/10.1093/bioinformatics/btr509 PMID: 21903627

74. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011; 7: 539. https://doi.org/10.1038/msb.2011.75 PMID: 21988835

75. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. Molecular Biology and Evolution. 2017; 34: 3299–3302. https://doi.org/10.1093/molbev/msx248 PMID: 29029172

76. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. Journal of Cheminformatics. 2011; 3: 33. https://doi.org/10.1186/1758-2946-3-33 PMID: 21982300

77. Kellogg EH, Leaver-Fay A, Baker D. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. Proteins. 2011; 79: 830–838. https://doi.org/10.1002/prot.22921 PMID: 21287615

78. Battey C. Cjbattey/Driftr: Driftr_V1.3 [Internet]. Zenodo; 2017. https://doi.org/10.5281/zenodo.345172

79. Ogino S, Gulley ML, den Dunnen JT, Wilson RB, Association for Molecular Patholpogy Training, Committtee E. Standard mutation nomenclature in molecular diagnostics: practical and educational challenges. J Mol Diagn. 2007; 9: 1–6. https://doi.org/10.2353/jmoldx.2007.060081 PMID: 17251329