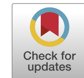
International Journal for Parasitology 48 (2018) 107–115

Contents lists available at ScienceDirect

International Journal for Parasitology

journal homepage: www.elsevier.com/locate/ijpara

Evolutionary epidemiology of schistosomiasis: linking parasite genetics with disease phenotype in humans



Tine Huyse [a,b,c,*,1], Nele A.M. Boon [a,b,1], Frederik Van den Broeck [d], Moustapha Mbow [e], Anurag Chaturvedi [a], Lynn Meurs [b], Filip A.M. Volckaert [a], Katja Polman [b]

[a] Laboratory of Biodiversity and Evolutionary Genomics, Biology, KU Leuven, Ch. de Bériotstraat 32, 3000 Leuven, Belgium
[b] Unit of Medical Helminthology, Department of Biomedical Sciences, Institute of Tropical Medicine, Nationalestraat 155, 2000 Antwerp, Belgium
[c] Department of Biology, Royal Museum for Central Africa, Leuvensesteenweg 13, 3080 Tervuren, Belgium
[d] Unit of Molecular Parasitology, Unit of Veterinary Protozoology, Department of Biomedical Sciences, Institute of Tropical Medicine, Nationalestraat 155, 2000 Antwerp, Belgium
[e] Department of Immunology, University Cheikh Anta Diop, Dakar, Senegal

A R T I C L E I N F O

Article history:
Received 26 January 2017
Received in revised form 15 June 2017
Accepted 5 July 2017
Available online 15 November 2017

Keywords:
Environmental association analysis
Molecular epidemiology
Schistosoma mansoni
Disease phenotype
cGMP-dependent protein kinase
Follistatin
Parasite fecundity

A B S T R A C T

Here we assess the role of parasite genetic variation in host disease phenotype in human schistosomiasis by implementing concepts and techniques from environmental association analysis in evolutionary epidemiology. Schistosomiasis is a tropical disease that affects more than 200 million people worldwide and is caused by parasitic flatworms belonging to the genus *Schistosoma*. While the role of host genetics has been extensively studied and demonstrated, nothing is yet known on the contribution of parasite genetic variation to host disease phenotype in human schistosomiasis. In this study microsatellite genotypes of 1561 *Schistosoma mansoni* larvae collected from 44 human hosts in Senegal were linked to host characteristics such as age, gender, infection intensity, liver and bladder morbidity by means of multivariate regression methods (on each parasite locus separately). This revealed a highly significant association between allelic variation at the parasite locus *L46951* and host infection intensity and bladder morbidity. Locus *L46951* is located in the 3′ untranslated region of the cGMP-dependent protein kinase gene that is expressed in reproductive organs of adult schistosome worms and appears to be linked to egg production. This putative link between parasite genetic variation and schistosomiasis disease phenotype sets the stage for further functional research.

© 2017 Australian Society for Parasitology. Published by Elsevier Ltd. All rights reserved.

1. Introduction

It is becoming increasingly feasible and affordable to generate molecular data for small-sized pathogens with the advent of new molecular techniques, thereby facilitating species diagnosis and epidemiological research. However, as pointed out by Restif (2009), a huge amount of molecular data is merely used in a diagnostic framework, and therefore left under-used. By integrating this molecular data with evolutionary theory and epidemiology (‘evolutionary epidemiology’; Archie et al., 2009) much more information can be extracted. ‘Environmental association analysis’, also called ‘genotype-environmental analysis’ (Lotterhos and Whitlock, 2015; Rellstab et al., 2015) is an emerging research field in evolutionary biology. This method identifies genetic variants associated with particular environmental factors and has the potential to uncover adaptive patterns that are not easily detected by traditional population genetic approaches (Rellstab et al., 2015). Here, we introduce concepts of environmental association analysis in evolutionary epidemiology, to link parasite genetic data with human epidemiological data on schistosomiasis in northern Senegal.

Schistosomiasis is a major poverty-related disease infecting more than 200 million people in developing countries, over 90% of whom live in Africa (Hotez and Kamath, 2009). The causative agents are blood flukes belonging to the genus *Schistosoma* (Digenea, Platyhelminthes). They have an obligate alternate life cycle in a freshwater snail host where asexual amplification generates thousands of clonal cercariae that infect the human host, in which sexual reproduction takes place. The most prevalent human schistosome species in Africa are *Schistosoma mansoni* and *Schistosoma haematobium*. Pathology is mainly caused by host immune responses against the eggs that are trapped in host tissues (Burke et al., 2009). Chronic infection with *S. mansoni* may result in liver fibrosis, spleen congestion and portal hypertension. The eggs of *S. haematobium* induce lesions in the vesical and ureteral walls or

* Corresponding author at: Department of Biology, Royal Museum for Central Africa, Leuvensesteenweg 13, 3080 Tervuren, Belgium. Fax: +32 32 45 75.
E-mail address: tine.huyse@africamuseum.be (T. Huyse).

[1] These authors contributed equally.

https://doi.org/10.1016/j.ijpara.2017.07.010
0020-7519/© 2017 Australian Society for Parasitology. Published by Elsevier Ltd. All rights reserved.

in the reproductive organs; chronic infection can result in fibrosis of the bladder and lower ureters (Gryseels et al., 2006).

Three commonly studied disease phenotypes of human schistosomiasis are the intensity of infection, the immune response to infection and schistosome-related pathology (Bethony and Quinnell, 2008). In a Brazilian pedigree study it was found that *S. mansoni* infection intensity, expressed as the number of eggs per gram of faeces (epg), was found to be highly over-dispersed, with only a small proportion of the human population displaying high infection intensities, often tied to certain households (Dessein et al., 1992). Bethony et al. (2002) showed that part of this variation in infection intensity could be explained by shared environment (household), accounting for 14% of the variance, and shared host genes (heritability), accounting for 21–27% of the variance. In the same study, a segregation analysis provided strong evidence for a major human co-dominant gene (SM1) controlling *S. mansoni* infection levels. Linkage analysis on a subset of the Brazilian subjects localised the major locus to chromosome 5, more specifically to the 5q31-q33 region (Marquet et al., 1996, 1999). This successful mapping of a gene region linked with infectious disease was also confirmed in a linkage study on a Senegalese population (Müller-Myhsok et al., 1997). The 5q31-q33 region contains several candidate loci involved in the regulation of the immune response to pathogens, such as genes coding for IL-13, IL-4, IL-5, IL-3 and IL-9, and the IFN regulatory factor-1 (Marquet et al., 1999). Subsequent candidate gene studies found associations between genetic variation in some of these genes, similar to the one coding for IL13, and infection intensity (Kouriba et al., 2005; Isnard et al., 2011; Grant et al., 2012) and resistance to *S. mansoni* re-infection after treatment (Gatlin et al., 2009).

Also, schistosome-related pathology is not random as it is more frequently found in certain families within a human population (Dessein et al., 1999). It appears, however, to be under different genetic control than infection intensity, since SM2, the major locus controlling liver fibrosis, is not linked to chromosome 5q31-33 (Bethony and Quinnell, 2008).

Whereas these studies showed that host genetic variation clearly affects infection intensity and schistosome-related pathology, much variation in disease phenotypes among hosts remains unexplained. For example, approximately 60–66% of the variation in infection intensity was unexplained in the above study by Dessein et al. (1992). Here we specifically look at the role of parasite genetic variation in host disease phenotype. Within spatiotemporal studies in northern Senegal we compiled several datasets on both parasite population genetics and host immuno-epidemiology. The parasite dataset consists of microsatellite genotypes of larval stages of *S. mansoni*, since adult schistosomes are inaccessible as they are located in the blood veins. Some microsatellite loci are in the untranslated regions (UTRs) of expressed sequence tags (ESTs). Because they are closely linked to coding regions in the genome, these EST-linked loci can be under selection and thus of particular interest to our study. The host dataset consists of disease phenotype information such as infection intensity and schistosome-related morbidity. We linked both datasets by applying a redundancy analysis (RDA), a technique used in environmental association analysis (Rao, 1964; Jombart et al., 2009; Rellstab et al., 2015). The specific aim was to explore the contribution of schistosome genetic variation to host infection intensity and morbidity caused by *S. mansoni* and *S. haematobium*.

2. Materials and methods

2.1. Study area and ethics statement

All data were previously collected during a large study on the epidemiology of schistosomiasis and innate immune responses (SCHISTOINIR: www.york.ac.uk/res/schistoinir), for which approval was obtained from the review board of the Institute of Tropical Medicine (Antwerp, Belgium), the ethical committee of the Antwerp University Hospital (Antwerp, Belgium) and 'Le Comité National d'Ethique de la Recherche en Santé' (Dakar, Senegal). Informed written consent was obtained from all participants. After the study, praziquantel (40 mg/kg) and mebendazole (500 mg) treatment were offered to all community members to treat and prevent schistosomiasis and soil-transmitted helminthiasis, respectively, according to WHO guidelines (WHO, 2006). All the data were anonymised prior to analysis.

Host demographic data (Meurs et al., 2012a,b; unpublished data) and parasite genotypes (Van den Broeck et al., 2014) were collected between February 2009 and January 2010 in northern Senegal in the villages of Pakh (16°24′9″N 15°48′42″W), Diokhor Tack (16°11′24″N 15°52′48″W) and Ndieumeul (also known as Thiekène, 16°13′12″N 15°51′36″W). The latter two are neighbouring villages along the western shore of Lac de Guiers, while Pakh is located a further 22 km north. The three villages are co-endemic for *S. mansoni* and *S. haematobium*. To our knowledge there have been no periodic anthelminthic programmes in Diokhor Tack or Ndieumeul (Meurs et al., 2012a,b); the community of Pakh received anthelminthic treatment in the framework of a study in 2006. Parasitological data on prevalence, infection intensity and morbidity were obtained from 599 individuals in Diokhor Tack, 99 individuals in Ndieumeul and 530 individuals in Pakh. Prevalences of *S. mansoni* infection, *S. haematobium* infection and mixed infection with both *Schistosoma* spp. were 55%, 44% and 32% in Diokhor Tack, 75%, 66% and 55% in Ndieumeul and 17%, 58% and 12% in Pakh, respectively. The prevalence of *S. mansoni*-specific hepatic fibrosis was 31% in Diokhor Tack, 19% in Ndieumeul and 8% in Pakh. The prevalence of *S. haematobium*-specific bladder morbidity was high, with 80% in Diokhor Tack and 91% in Ndieumeul (Meurs et al., 2012a,b). In Pakh 17% of the individuals showed bladder morbidity (unpublished data).

2.2. Host characteristics and disease phenotype

Host disease phenotype was defined using four measurements: (i) *S. mansoni* infection intensity, (ii) *S. haematobium* infection intensity, (iii) *S. mansoni*-specific lesions of the liver and (iv) *S. haematobium*-specific lesions of the bladder (Tables 1 and 2).

In short, two stool and two urine samples were collected from all participants. *Schistosoma mansoni* infection intensity was determined by duplicate 25 mg Kato-Katz examinations on each stool sample (Katz et al., 1972; Polderman et al., 1985) and expressed as the number of eggs per gram of feces (epg). *Schistosoma haematobium* infection was determined after urine filtration using a filter of 12 μm pore size (Isopore, USA) and expressed as the number of eggs per 10 ml of urine (ep10ml). Pathological lesions associated with schistosome infection were recorded by ultrasound according to the Niamey guidelines (Richter et al., 1996). The urinary bladder score (UBS score) is a measure for *S. haematobium*-specific bladder morbidity, while the liver image pattern (liver IP) is a measure of *S. mansoni*-specific morbidity of the liver. In addition, circulating anodic antigen (CAA) concentrations were determined from one serum sample per individual as described by Corstjens et al. (2014) using a random subset of the participants ($n = 19$). CAA, a *Schistosoma*-specific, gut-derived glycoprotein steadily excreted by metabolically active adult worms, can be interpreted as a measure of current worm burden in humans (Deelder et al., 1994; Agnew et al., 1995; Polman et al., 2001). Regarding CAA levels and egg counts as a reflection of worm burden and egg production, respectively, a proxy of fecundity can be introduced, which was calculated as follows: epg/picogram of CAA per ml of plasma (Wilson et al., 2014).

Table 1
Descriptors of categorical host characteristics: gender, village, liver morbidity (IP score), bladder morbidity (UBS score) and mixed infection with *Schistosoma haematobium*.

Parameter	*n* (total = 44)	%
Gender		
Male	16	36.4
Female	28	63.6
Village		
Pakh (code 200)	6	13.6
Ndieumeul (code 501)	29	65.9
Diokhor Takh (code 502)	8	18.2
Liver morbidity (liver IP)		
No morbidity (0–1)	34	77.3
Clear morbidity (2–4)	8	18.2
NA	2	4.5
Bladder morbidity (UBS score)		
No morbidity (0–1)	21	47.7
Clear morbidity (2–4)	21	47.7
NA	2	4.5
Mixed infection (with *S. haematobium*)		
No (= single *S. mansoni* infection)	11	25.0
Yes	31	70.5
NA	2	4.5

UBS scores of 0 or 1 were considered as urinary schistosomiasis being unlikely and treated as category '0' in further analysis. UBS scores ≥2 are considered as *S. haematobium*-specific bladder lesions, category '1'. Liver IP scores lower than 2 were considered as intestinal schistosomiasis being unlikely and treated as category '0'; scores ≥2 indicate *Schistosoma mansoni*-specific liver morbidity and treated as category '1'. NA, not available.

Other host characteristics that were defined were gender, age, village, and single *S. mansoni* infection and mixed infection with *S. haematobium* (Tables 1 and 2). CAA concentrations, epg, ep10ml and fecundity (epg/CAA) were log-transformed during all analyses to comply with model assumptions. Age was included in models as a categorical variable (four age groups, see Table 2).

2.3. Parasite genetic data: S. mansoni genotyping

Schistosoma mansoni eggs were filtered from positive stool samples using a Pitchford and Visser funnel (Pitchford and Visser, 1975), subsequently concentrated and hatched in bottled mineral water. Using a binocular microscope, individual miracidia were pipetted in 3 μl per miracidium onto Whatman FTA® classic cards. In Belgium, 3 mm discs containing each miracidium were excised from the cards and DNA was extracted using the Nucleospin Tissue kit (Macherey Nagel, Belgium) following the manufacturer's instructions. Each individual parasite (1561 *S. mansoni* larvae) was genotyped with nine microsatellite loci combined in a single multiplex: *L46951*, *smd25*, *smd28* and *smd89* (Durand et al., 2000); *CA11-1* and *S9-1* (Blair et al., 2001); *smd11*, *smd43* and *sdma28* (Curtis et al., 2001).

Basic Local Alignment Tool (BLAST; https://blast.ncbi.nlm.nih.gov) results for the primer sequences indicated that two of these loci are located in regions that are expressed (EST-linked): *L46951* is located in the 3′ UTR region of a cGMP-dependent protein kinase gene (cGK, GenBank Accession No. **Smp_123290**), and *CA11-1* is located in a B cell receptor-associated protein mRNA (GenBank Accession No. DQ480542). *SMD28* is linked to a putative gene but its function is not known (GenBank Accession No. **Smp_138400**).

2.4. Statistical analysis

2.4.1. Associations between S. mansoni genetic data and host characteristics

The flow diagram (Fig. 1) provides an overview of statistical analyses. First, the human host characteristics were analysed separately. Epidemiological studies suggest that host characteristics are associated in endemic schistosomiasis settings. To test associations between host characteristics in this particular dataset, two multiple linear regression models (MLR) were made with *S. haematobium* infection intensity or *S. mansoni* infection intensity as a dependent variable, and all other host characteristics (host gender, age, bladder morbidity, liver morbidity, village) as independent variables. This was also performed using parasite fecundity as a dependent variable, for a subset of 19 hosts.

To explore associations between the parasite genetic data and host characteristics, a RDA (Rao 1964) was performed. RDA is a constrained ordination method based on Euclidean distances and combines a multivariate regression with a principal component analysis of the table of fitted values (Borcard et al., 2011). The use of RDA to associate parasite genetic data with host characteristics ("environmental variables") does not rely on an underlying genetic model such as Hardy–Weinberg equilibrium, while theoretical population genetic approaches often do (Jombart et al., 2009, Joost et al., 2013). This is relevant for schistosomes, considering their complex life cycle which includes an asexual reproduction step within a snail host. It also relies less on overall levels of genetic differentiation among populations than outlier detection methods (Hancock et al., 2010) and is therefore suitable for schistosome populations as they often show a high diversity within a host and low levels of genetic differentiation among hosts (Curtis and Minchella, 2000; Van den Broeck et al., 2014). Allele frequencies of the offspring (miracidia) were grouped per host individual, hereafter referred to as infrapopulations. These infrapopulations, grouped per locus, were used to represent the parasite genetic variation (the dependent variables). Host characteristics represented the independent (i.e. constraining) variables. Each locus was considered the unit of analysis because the inclusion of all loci together may average out the correlation of allele frequencies with host characteristics at a single locus. Before performing the RDA analysis, the allele frequencies were centred but not scaled. Nine separate dependent allelic matrices, one for each locus, were tested against one and the same independent matrix with host characteristics, and thus, respectively, nine separate RDA models were built. To test for significance of the model, and to test for the significance of the marginal effects of each constraining variable separately, an

Table 2
Schistosoma mansoni and *Schistosoma haematobium* infection intensities per age group for the complete dataset.

		S. mansoni infection intensity (epg)		*S. haematobium* infection intensity (ep10ml)	
Age (years)	No.	Median	Range	Median	Range
<9	12	435	20–4140	36.0	0–1111
10–14	13	770	370–4360	29.5	0–217
15–19	5	850	110–1010	46.7	0–91
>20	14	560	10–6470	1.0	0–17
ALL	44	560	10–6470	4.5	0–1111

epg, number of eggs per gram of faeces; ep10ml, number of eggs per 10 ml of urine.

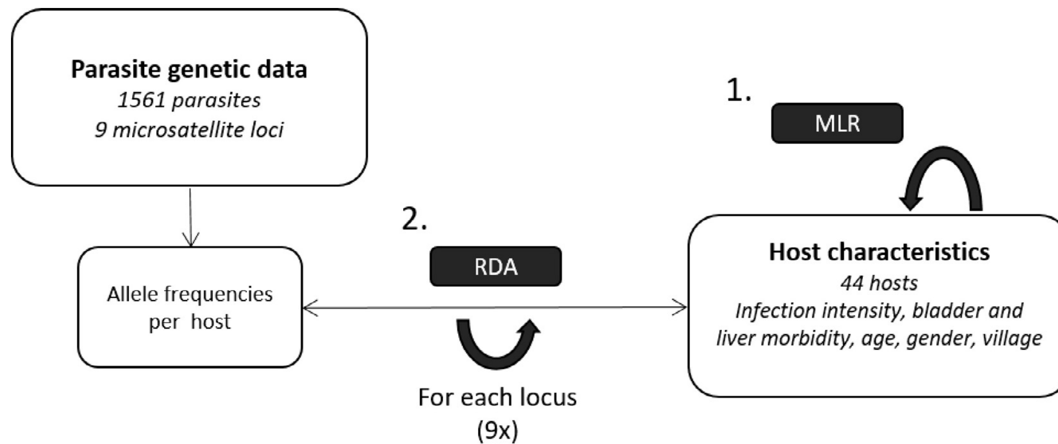


Fig. 1. Exploring associations between the parasite genetic data and host characteristics. This flowchart summarises the association of the parasite genetics dataset with the dataset of host characteristics. First the independence of variables within the host dataset was investigated (1) by performing multivariate linear regressions (MLR) among host characteristics. To link parasite genetics with host characteristics (2), the microsatellite genotypes of 1561 individual parasites (miracidia) were converted to allele frequencies per host individual. A redundancy analysis (RDA) was used to correlate the allele frequencies of 44 parasite populations with host characteristics. RDA was applied nine times, for each parasite locus separately, with the same dataset of host variables.

ANOVA-like permutation test of 10,000 permutations was performed, as available in the vegan package (Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H., 2013. Package "vegan." R Package. ver. 2.0–8.). Significant RDA models are represented as a Venn-diagram, which shows the result of a variation partitioning of the parasite genetic variables by different independent variables (Borcard et al., 1992). To correct for multiple testing during RDA analysis, sequential Bonferroni correction was used. To confirm the results on offspring genotypes, all analyses were repeated using predicted parental allele frequencies. Since we only have access to the offspring population, parental genotypes were reconstructed based on offspring genotypes through a full-pedigree likelihood method of Jones and Wang (2010) implemented in COLONY software (see Supplementary Table S1).

RDA was performed using the R-packages adegenet (Jombart, 2008), ade4 (Dray and Dufour, 2007) and vegan (Oksanen et al., 2013, cited earlier). R version 2.13.2 (R Development Core Team, 2015) was used for all statistical analyses.

2.4.2. Association of locus L46951 with host characteristics

The parasite locus that we observed to be significantly associated with host characteristics in RDA (i.e. locus L46951; see Section 3.2) was analysed in more detail. Firstly, hypothesis error plots (HEplots) were constructed to visualise how predictors (= alleles) relate collectively to the response variables (= host characteristics), based on multivariate regression analysis with the R-package heplots (Friendly 2007; Supplementary Fig. S1). Secondly, the association between alleles of this locus and parasite fecundity was tested on the sub-dataset (n = 19). Allelic frequencies of locus L46951 were regressed separately over parasite fecundity by using a univariate linear regression (ULRA).

2.4.3. Data accessibility

Microsatellite genotypes and host phenotype data are deposited in the DRYAD databank, doi: 10.5061/dryad.t4s0c.

3. Results

3.1. Host characteristics, disease phenotype and S. mansoni genetic data

Table 1 gives an overview of all categorical host characteristics (gender, village, liver morbidity, bladder morbidity, single and mixed infections). Note that most individuals (70.5%) had a mixed infection with *S. haematobium* and *S. mansoni*. For this reason, further analyses include both mixed and single infections, as separate analyses per infection category were not possible due to the low number of individuals with single infections (n = 11). Table 2 describes the infection intensity of *S. mansoni* and *S. haematobium* across the four age groups. *Schistosoma mansoni* infection intensities, CAA measurements and parasite fecundity per age group are shown in Supplementary Table S2. In total 2891 miracidia collected from 63 host individuals were processed for genotyping at all nine microsatellite loci.

3.2. Associations between S. mansoni genetic data and host characteristics

In order to ensure high data quality we first selected 1692 parasites with at least five out of the nine loci successfully amplified. We then retained only those infrapopulations with a minimum of 18 parasites, resulting in a dataset of 1561 parasites from 44 host individuals. The sizes of infrapopulations ranged from 18 to 86 with a median value of 31. The threshold of 18 was chosen, as after this value the rarefaction curve reached a plateau (data not shown), suggesting that most of the variation present in the parasite population is covered. The results of the two linear regression models that test for associations between *S. mansoni* or *S. haematobium* infection intensity and host characteristics were as follows: *S. haematobium* infection intensities were significantly correlated with age (four age groups, Table 2) (P = 0.009, R^2 = 20.3%); individuals older than 20 years had a lower *S. haematobium* infection intensity compared with children younger than 9 years of age (Table 2). *Schistosoma mansoni* infection intensity was not associated with host age (P = 0.34, R^2 = 1.1%) or with other host characteristics. Table 2 describes *S. mansoni* and *S. haematobium* infection intensities per age group. Parasite fecundity was positively associated with *S. mansoni* infection intensity (P < 0.001, R^2 = 45,5%) and negatively associated with CAA concentrations (P < 0.001, R^2 = 62.9%).

RDA analyses revealed a significant relationship between *S. mansoni* genetic variation at locus *L46951* and *S. mansoni* infection intensity, host age and bladder morbidity. Allelic variation at this locus was partitioned over these three host characteristics. Detailed information concerning the seven corresponding RDA models is shown in Table 3 and Fig. 2. The RDA model with the three variables included (Table 3, model 7) was highly significant

Table 3
Summary statistics for the redundancy analysis (RDA) on the allelic variation of locus *L46951* in the parasite population (miracidia) and the characteristics of the associated host. The *P* value of the overall RDA models and marginal terms of the model are based on 10,000 simulations. The information of these seven RDA models is summarised in a Venn diagram (Fig. 2) that represents variation partitioning of the parasite allelic variation at locus *L46951* by three host variables.

Locus *L46951*	Constrained variables	R^2 adj	*P* value model	*P* value marginal terms
Model 1	*S. mansoni* infection	11.96%	0.0024	/
Model 2	Age	6.97%	0.0568	/
Model 3	Bladder morbidity	6.97%	0.0205	
Model 4	*S. mansoni* infection Age	16.29%	0.0047	0.0061
				0.1186
Model 5	Bladder morbidity	17.28%	0.0007	0.0120
	S. mansoni infection			0.0045
Model 6	Bladder morbidity	13.00%	0.0147	0.0227
	Age			0.0771
Model 7	*S. mansoni* infection	21.11%	0.0029	0.0102
	Age			0.1315
	Bladder morbidity			0.0125

R^2 adj, adjusted R^2, variance in allelic frequencies of locus *L46951* explained by the constrained variables.

($P = 0.0029$) and the adjusted R^2 (R^2_{adj}) indicated that 21.1% of the total genetic variation (~constrained variation) at locus L46951 was explained by these constrained variables (unadjusted R^2 = 31.0%). No confounding effects of mixed infection with *S. haematobium* or the size of the parasite infrapopulation were found. Variation partitioning of this model indicated that most of the genetic variation was related to *S. mansoni* infection intensity (8.1%). Age and bladder morbidity accounted for, respectively, 3.8% and 4.8% of the genetic variation (Table 3, Fig. 2). A RDA model (model 1, Table 3) with only *S. mansoni* infection intensity explained 12% of the genetic variation at this locus ($P = 0.0024$).

The three most abundant alleles of locus *L46951* were 169, 172 and 175, with a mean allelic frequency per host of 43.4%, 23.2% and 26.2%, respectively. Both allele ^{172}*L46951* ($P < 0.0001$, $R^2 = 0.48$) and allele ^{175}*L46951* ($P = 0.008$, $R^2 = 0.14$) were negatively correlated with allele ^{169}*L46951.* The other four alleles (163, 166, 178 and 181) were rare with mean allelic frequencies per host of 0.1%, 5.9%, 0.7% and 0.3%, respectively.

In RDA model 7 (Table 3), the highest scores for alleles correlating with the first constrained axis (RDA1, explaining 89.4% of the total constrained variation) were allele ^{169}*L46951* (score = 0.31), allele ^{172}*L46951* (score = -0.39) and allele ^{175}*L46951* (score = 0.06). *Schistosoma mansoni* infection intensity and bladder morbidity were positively correlated with RDA1 and consequently positively related to allele ^{169}*L46951* and negatively to allele ^{172}*L46951*. Age was positively correlated with RDA1 until the age of 20. The second constrained axis (RDA2, 8.6% of constrained variation) was mainly correlated with allele ^{175}*L46951* (score = 0.12). Bladder morbidity (score = 0.79) and *S. mansoni* infection intensity (score = -0.52) were the host variables that correlated most strongly with RDA2. This indicates that allele ^{175}*L46951* correlates negatively with *S. mansoni* infection intensity. Fig. 3 shows a scatterplot and the results of a univariate regression of allele ^{169}*L46951* with *S. mansoni* infection intensity. The pattern was opposite for allele ^{172}*L46951*; this allele was negatively correlated with bladder morbidity (P = 0.012) while no relationship was found with ^{175}*L46951* (see Supplementary Fig. S2). The same results were found when using a higher or lower threshold for the infrapopulation size (data not shown). A significant association between ^{169}*L46951* and *S. mansoni* infection intensity was also found when predicted parental allelic frequencies were used (see Supplementary Table S1).

3.3. Association of locus L46951 with host characteristics

The relationships of alleles ^{169}*L4695* and ^{172}*L46951* with host characteristics were highly significant in the multivariate regression (see Supplementary Fig. S1, Supplementary Table S3). The regression of the frequency of allele ^{169}*L46951* over fecundity was significant (Fig. 4). Parasite fecundity accounted for 34.3% of the variation and no confounding variables were found. In contrast to the positive correlation with allele ^{169}*L46951*, the correlation with alleles ^{172}*L46951* or ^{175}*L46951* as a dependent variable was negative (rc = −0.03, $P = 0.03$, R^2 = 19.5% for allele ^{172}*L46951* and r c = −0.02, $P = 0.008$, R^2 = 30.3% for allele ^{175}*L46951*).

4. Discussion

To our knowledge, this is the first study showing a significant association between helminth parasite genetic variation and host disease phenotype. We were able to link the presence of specific alleles in *S. mansoni* parasite populations, isolated from human hosts, with specific disease phenotypes. Allele ^{169}*L46951* was positively associated with infection intensity, while ^{172}*L46951* was negatively associated with infection intensity. It was shown that 21% of the total *S. mansoni* allelic variation at microsatellite locus *L46951* was explained by infection intensity, host age and bladder morbidity, with infection intensity providing the highest contribution (12%). The explained variation of locus *L46951* in our study was high compared with other environmental association studies that used RDA to associate population genetic variation with environmental variables (eg. James et al., 2011; Vangestel et al., 2012; Orsini et al., 2013). This is particularly remarkable when considering (i) the small geographical scale and (ii) the single-locus approach in this study (James et al., 2011; Vangestel et al., 2012). In addition, three single explanatory variables (infection intensity, host age and bladder morbidity) contributed to a significant model.

Previous studies tried to link schistosome genetic variation with disease outcome. Brouwer et al. (2003) found genetic differences, measured by randomly amplified polymorphic DNA (RAPD), between *S. haematobium* populations from children with varying pathology. However, they could not link pathology to specific parasite genotypes or genes due to the limitations of the RAPD technique. Gasmelseed et al. (2014) also used a RAPD approach in a Sudanese setting but could not confirm the results of Brouwer and colleagues (2003). Barbosa et al. (2013) used 15 microsatellite markers and found geographic clustering in *S. mansoni* over a scale of 8 km, but could not link this with demographic or epidemiological host characteristics. However, schistosome eggs were pooled per host sample and all microsatellite loci were summarised in one differentiation parameter, which may decrease the power to detect associations between single parasite alleles and host variables.

The environmental association analysis used in this study, on individually genotyped parasite larvae and using a single-locus

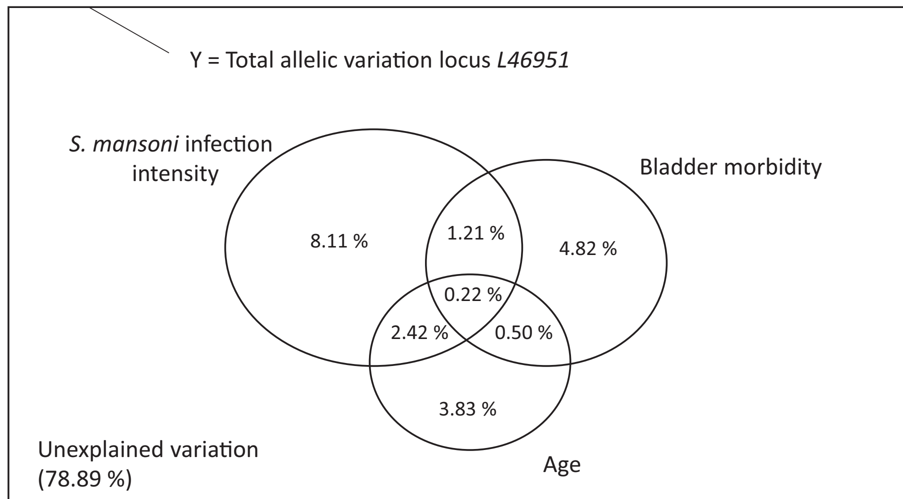


Fig. 2. Venn diagram of the variation partitioning of the response data set, representing the allelic frequencies of locus *L46951* among the variables: *Schistosoma mansoni* infection intensity, bladder morbidity, and age. The variation partitioning is based on the adjusted R^2-values from the seven redundancy analysis-models, represented in Table 3.

approach, has proven to be a powerful tool to uncover subtle patterns in parasite genetic variation that might be missed by traditional methods for the detection of outlier loci, based on population genetic differentiation (Rellstab et al., 2015). Also, the possibility to construct models with several (environmental/host) variables simultaneously is especially relevant in the case of complex disease phenotypes such as schistosomiasis.

The present study contained 44 infrapopulations, of which only a few host individuals had low infection intensities. Fig. 3 gives the impression that the two lower infection intensity points are mainly responsible for the positive trend. However, when we removed these two lower infection points, the positive association between allele ^{169}*L46951* and infection intensity remained (rc = 0.03569, P = 0.006, R^2 = 15.7%). We are therefore confident that this association is statistically robust. Future studies with larger sample sizes, including more host individuals from lower infection classes, would however be desirable.

This study shows that the neutrality of microsatellite markers should not be taken for granted. Adaptive loci are not suitable as a tool for population genetic studies and detection of demographic patterns (Selkoe and Toonen, 2006). The neutrality of the loci should therefore be tested prior to their use in population genetic studies.

After a BLAST search against the *S. mansoni* complete genome, *L46951* appeared to be a trinucleotide (GAA) repeat that occurs in the 3′ UTR region of the cGMP-dependent protein kinase (cGK) gene (**Smp_123290**; Leutner et al., 2011), which is located on chromosome 6 (Zerlotini et al., 2013). Follistatin (**Sm_123300**; Leutner et al., 2013) and the chromodomain helicase DNA binding protein (**Sm_123320**; Zerlotini et al., 2013) are located downstream of *L46951*. Interestingly, both the cGK gene and follistatin have been studied in *S. mansoni*. The cGK belongs to the family of serine/threonine-specific protein kinases that are present in a variety of eukaroytes which play a role in signalling cascades of muscle cells (Schlossmann et al., 2003; Hofmann et al., 2009). The gene was recently studied by Leutner et al. (2011) who identified six possible cGK homologs in the *S. mansoni* genome. One of these predictions, *Smp_123290*, which contains the *L46951* locus, was cloned and characterised. Real-time PCR and *in situ* hybridisation experiments revealed a gonad-preferential expression profile in both genders; transcriptional activity was mainly detected in the ovary, the vitellarium, and testes of adult worms. Treatment of adult schistosomes *in vitro* with the inhibitory cGMP analogue resulted in slow motion, oocyte congestion, and a reduction in egg production by 30% (Leutner et al., 2011). Follistatin, located 32,016 bp from locus *L46951*, is a regulator of the TGFβ signalling pathway that plays a role in regulating mitotic activity and egg production in paired *S. mansoni* females (Leutner et al., 2013).

Since we can only genotype the offspring population (miracidial larvae isolated from stool samples), we reconstructed parental genotypes and used these in the RDA analysis. The same positive association between the frequency of ^{169}*L46951* and infection intensity was found. This suggests that adult parasite populations with allele ^{169}*L46951* have more offspring than those without this allele. Even though we can only measure eggs that are excreted (a portion of the eggs remains stuck in the host tissue), this might suggest that those adults with allele ^{169}*L46951* produce more eggs and thus have a higher fecundity than those without this allele. This was corroborated by the use of a proxy for fecundity for a subset of the cohort (n = 19), which was strongly positively associated with allele ^{169}*L46951* (Fig. 4). A larger dataset would allow testing of the robustness of the significant association.

An accurate quantification of infection intensity is very important in these analyses. Here we performed duplicate Kato-Katz examinations on two different stool samples, allowing a sufficiently sensitive and accurate quantification of schistosome egg excretion in our study setting (Utzinger et al., 2015). In addition, we measured CAA concentrations in the sera of a small subsample of participants. Serum CAA levels have been reported to correlate with schistosome worm numbers in experimental animals, thereby providing a valuable alternative measure of infection intensity. For this reason, it would be good to include CAA (or CCA) detection tests in future evolutionary epidemiologic studies.

Both the location of the microsatellite marker (in a cGK gene and nearby follistatin) and the link between genetic variation at this locus with infection intensity and with (a proxy for) parasite fecundity, suggests that ^{169}*L46951* is linked with *S. mansoni* parasite fecundity. Also other studies have suggested a link between protein kinases and schistosome reproduction (Morel et al., 2014). Eggs that are trapped in host tissues are the main cause of pathology, so vaccines or drugs that interact with parasite reproduction are therefore considered as a potential tool for control (Wilson et al., 2014). Accordingly, protein kinases are new important drug targets in schistosomes (Morel et al., 2014; Walker et al., 2014).

In addition to the link with *S. mansoni* infection intensity, there was also a significant link between *S. mansoni* genetic variation and

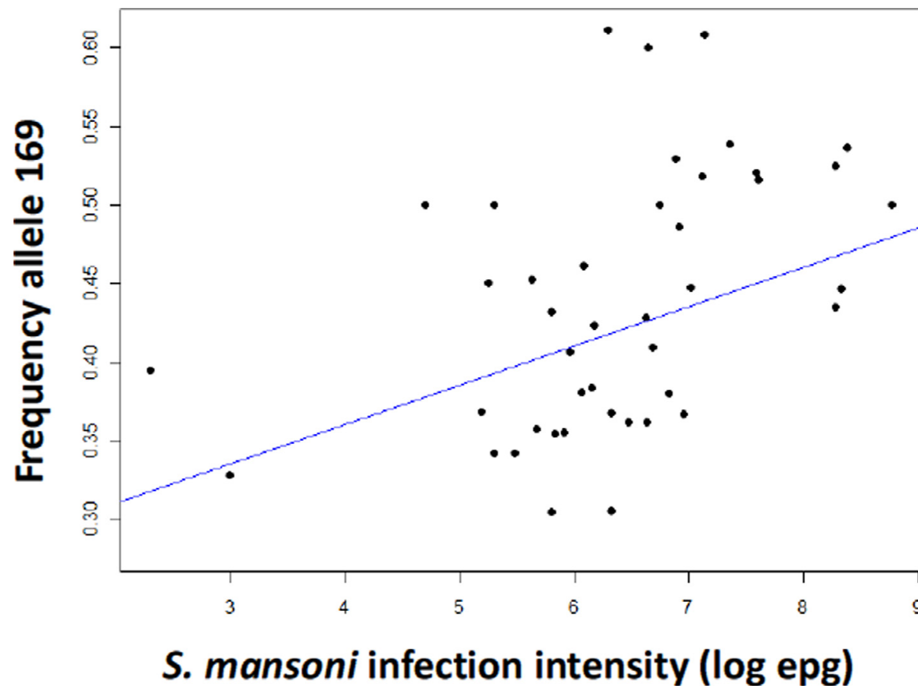


Fig. 3. Scatter plot of the frequency of allele [169]*L46951* over *Schistosoma mansoni* infection intensity (log eggs per gram of faeces (epg)). The univariate regression model was highly significant ($P = 0.003$, $R^2 = 19.38\%$).

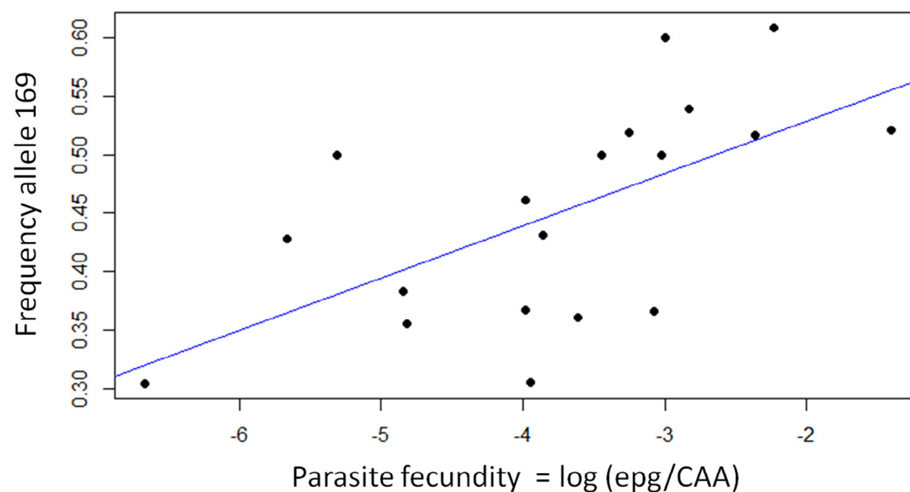


Fig. 4. Scatter plot of the frequency of allele [169]*L46951* over parasite fecundity (log (eggs per gram of faeces/circulating anodic antigen), on a subset of 19 observations. The positive regression coefficient was highly significant ($P = 0.005$, $R^2 = 34.4\%$).

host bladder morbidity and age. This was not due to multicollinearity between these variables and infection intensity, as variation partitioning clearly showed the statistically independent association of both variables with the allelic frequencies of locus *L46951*. The association between *S. mansoni* genetic variation and bladder morbidity is at first sight counter-intuitive as *S. mansoni* eggs are usually excreted via the stool. However, mixed *S. mansoni* and *S. haematobium* infections are very common in northern Senegal and this can lead to heterologous pairing between both species (Southgate et al., 1998). If *S. mansoni* females mate with *S. haematobium* males this will result in *S. mansoni* eggs in urine because the male defines the egg-laying site (Brant and Loker, 2005). In northern Senegal, Ernould et al. (1999) found that 31% of urine samples contained ectopic *S. mansoni* eggs while Meurs et al. (2012b) found that 13% of urine samples from people with mixed infections had ectopic *S. mansoni* eggs. The latter study showed that *S. haematobium*-infected people eliminating *S. mansoni* eggs both via urine and via stool had the highest risk of bladder morbidity.

Further experiments are needed to make any conclusion about the causality of the observed association between the parasite locus and host infection intensity. Secondly, the functional mechanism of the association needs to be disentangled. As discussed above, allelic variation of locus *L46951* may be linked with mutations upstream in the coding region of the cGK gene or with other genes downstream ('hitch-hiking'). Therefore, future studies should investigate this association at the level of the entire chromosome or even genome to determine the polygenic nature of the trait. Full exome analysis on single larval schistosome stages is possible nowadays but still expensive (Chevalier et al., 2014).

Another strategy is transcriptomic analysis (RNA-seq) to understand the gene expression changes and study gene regulatory networks. Both genomic and transcriptomic approaches allow a focus on molecular pathways (such as, for instance, the TGFβ signalling pathway) whose concerted action is much more likely to reflect the association with the host phenotype than single genes (Jackson, 2015). To perform RNAseq, adult worms and thus an experimental set-up are needed. Finally, microsatellite polymorphisms in 3′ UTR regions themselves can influence gene expression. For example, polymorphism can influence the binding of miRNA in the 3′ UTR region (Clop et al., 2006; Hon and Zhang, 2007; Bhattacharya et al., 2014; Chaturvedi et al., 2014) or alter chromatin structure, which is known to modulate DNA accessibility to DNA binding proteins such as DNA polymerase and thus affect gene expression (Xi et al., 2010).

Our results suggest parasite genetic variation to be an important factor in explaining the variation in host disease phenotype in human schistosomiasis. Further research, preferably with a larger sample size, is needed to test whether our findings are replicable and consistent across space and time. This study highlights the potential of environmental association analysis in uncovering subtle patterns in parasite genetic variation that may not be picked up with traditional population genetic approaches. The incorporation of ecological concepts and principles in the field of parasitic diseases may be a promising tool in ongoing or future molecular epidemiological studies.

Acknowledgements

We thank the field team of Richard Toll, A. Yague, M. Diop, M. Wade and N. Sy for assistance with sample collection and microscopic analysis. We also thank Hanne Dekort and Pascal Hablützel for useful comments on an earlier version of this manuscript and two anonymous referees for providing useful suggestions. Govert van Dam is acknowledged for the CAA measurements. Finally, special thanks to the people in Ndieumeul, Diokhor Tack and Pakh, Senegal, for their kind participation in the study. NB and FVdB benefited from a VLADOC PhD fellowship of the Flemish Interuniversity Council (VLIR-UOS), Belgium. Research was funded with a Research Grant of the Research Foundation—Flanders, Belgium (contract G.0552.10), a Krediet to navorsers (contract 1516913N) to TH and from travel grants awarded by VLIR-UOS to FVDB and NB.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ijpara.2017.07.010.

References

Agnew, A., Fulford, A.J., De Jonge, N., Krijger, F.W., Rodriguez-Chacon, M., Gutsmann, V., Deelder, A.M., 1995. The relationship between worm burden and levels of a circulating antigen (CAA) of five species of *Schistosoma* in mice. Parasitology 111, 67–76.

Archie, E., Luikart, G., Ezenwa, V.O., 2009. Infecting epidemiology with genetics: a new frontier in disease ecology. Trends Ecol. Evol. 24, 21–30. https://doi.org/10.1016/j.tree.2008.08.008.

Barbosa, L.M., Silva, L.K., Reis, E.A., Azevedo, T.M., Costa, J.M., Blank, W.A., Reis, M.G., Blanton, R.E., 2013. Characteristics of the human host have little influence on which local *Schistosoma mansoni* populations are acquired. PLoS Negl. Trop. Dis. 7, e2572. https://doi.org/10.1371/journal.pntd.0002572.

Bethony, J., Williams, J.T., Blangero, J., Kloos, H., Gazzinelli, A., Soares-Filho, B., Coelho, L., Alves-Fraga, L., Williams-Blangero, S., Loverde, P.T., Correa-Oliveira, R., 2002. Additive host genetic factors influence fecal egg excretion rates during *Schistosoma mansoni* infection in a rural area in Brazil. Am. J. Trop. Med. Hyg. 67, 336–343.

Bethony, J.M., Quinnell, R.J., 2008. Genetic epidemiology of human schistosomiasis in Brazil. Acta Trop. 108, 166–174. https://doi.org/10.1016/j.actatropica.2007.11.008.

Bhattacharya, A., Ziebarth, J.D., Cui, Y., 2014. PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. Nucleic Acids Res. 42, D86–D91. https://doi.org/10.1093/nar/gkt1028.

Blair, L., Webster, J., Barker, G., 2001. Isolation and characterization of polymorphic microsatellite markers in *Schistosoma mansoni* from Africa. Mol. Ecol. Notes c, 93–95.

Borcard, D., Gillet, F., Legendre, P., 2011. Numerical Ecology with R. Springer, New York, NY, USA.

Borcard, D., Legendre, P., Drapeau, P., 1992. Partialling out the spatial component of ecological variation. Ecology 73, 1045–1055. https://doi.org/10.2307/1940179.

Brant, S.V., Loker, E.S., 2005. Can specialized pathogens colonize distantly related hosts? Schistosome evolution as a case study. PLoS Pathog. 1, 0167–0169. https://doi.org/10.1371/journal.ppat.0010038.

Brouwer, K.C., Ndhlovu, P.D., Wagatsuma, Y., Munatsi, A., Shiff, C.J., 2003. Urinary tract pathology attributed to *Schistosoma haematobium*: does parasite genetics play a role? Am. J. Trop. Med. Hyg. 68, 456–462.

Burke, M.L., Jones, M.K., Gobert, G.N., Li, Y.S., Ellis, M.K., McManus, D.P., 2009. Immunopathogenesis of human schistosomiasis. Parasite Immunol. 31, 163–176. https://doi.org/10.1111/j.1365-3024.2009.01098.x.

Chaturvedi, A., Raeymaekers, J.A.M., Volckaert, F.A.M., 2014. Computational identification of miRNAs, their targets and functions in three-spined stickleback (*Gasterosteus aculeatus*). Mol. Ecol. Resour. 14, 768–777. https://doi.org/10.1111/1755-0998.12223.

Chevalier, F.D., Valentim, C.L.L., LoVerde, P.T., Anderson, T.J.C., 2014. Efficient linkage mapping using exome capture and extreme QTL in schistosome parasites. BMC Genomics 15, 617, pmid:25048426.

Clop, A., Marcq, F., Takeda, H., Pirottin, D., Tordoir, X., Bibé, B., Bouix, J., Caiment, F., Elsen, J.-M., Eychenne, F., Larzul, C., Laville, E., Meish, F., Milenkovic, D., Tobin, J., Charlier, C., Georges, M., 2006. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. Nat. Genet. 38, 813–818. https://doi.org/10.1038/ng1810.

Corstjens, P.L.A.M., De Dood, C.J., Kornelis, D., Fat, E.M.T.K., Wilson, R.A., Kariuki, T. M., Nyakundi, R.K., Loverde, P.T., Abrams, W.R., Tanke, H.J., Van Lieshout, L., Deelder, A.M., Van Dam, G.J., 2014. Tools for diagnosis, monitoring and screening of *Schistosoma* infections utilizing lateral-flow based assays and upconverting phosphor labels. Parasitology 141, 1841–1855. https://doi.org/10.1017/S0031182014000626.

Curtis, J., Minchella, D.J., 2000. Schistosome population genetic structure: when clumping worms is not just splitting hairs. Parasitol. Today 16, 68–71.

Curtis, J., Sorensen, R.E., Page, L.K., Minchella, D.J., 2001. Microsatellite loci in the human blood fluke *Schistosoma mansoni* and their utility for other schistosome species. Mol. Ecol. Notes 1, 143–145. https://doi.org/10.1046/j.1471-8278.2001.00053.x.

Deelder, A.M., Qian, Z.L., Kremsner, P.G., Acosta, L., Rabello, A.L., Enyong, P., Simarro, P.P., van Etten, E.C., Krijger, F.W., Rotmans, J.P., 1994. Quantitative diagnosis of *Schistosoma* infections by measurement of circulating antigens in serum and urine. Trop. Geogr. Med. 46, 233–238.

Dessein, A.J., Couissinier, P., Demeure, C., Rihet, P., Kohlstaedt, S., Carneiro-Carvalho, D., Ouattara, M., Goudot-Crozel, V., Dessein, H., Bourgois, A., 1992. Environmental, genetic and immunological factors in human resistance to *Schistosoma mansoni*. Immunol. Invest. 21, 423–453.

Dessein, A.J., Hillaire, D., Elwali, N.E., Marquet, S., Mohamed-Ali, Q., Mirghani, A., Henri, S., Abdelhameed, A.A., Saeed, O.K., Magzoub, M.M., Abel, L., 1999. Severe hepatic fibrosis in *Schistosoma mansoni* infection is controlled by a major locus that is closely linked to the interferon-gamma receptor gene. Am. J. Hum. Genet. 65, 709–721. https://doi.org/10.1086/302526.

Dray, S., Dufour, A.B., 2007. The ade4 package: implementing the duality diagram for ecologists. J. Stat. Softw. 22, 1–20. 10.1.1.177.8850.

Durand, P., Sire, C., Théron, A., 2000. Isolation of microsatellite markers in the digenetic trematode *Schistosoma mansoni* from Guadeloupe island. Mol. Ecol. 9, 997–998.

Ernould, J.C., Ba, K., Sellin, B., 1999. Increase of intestinal schistosomiasis after praziquantel treatment in a *Schistosoma haematobium* and *Schistosoma mansoni* mixed focus. Acta Trop 73, 143–152.

Friendly, M., 2007. HE Plots for multivariate linear models. J. Comput. Graph. Stat. 16, 421–444. https://doi.org/10.1198/106186007X208407.

Gasmelseed, N., Karamino, N.E., Abdelwahed, M.O., Hamdoun, A.O., Elmadani, A.E., 2014. Genetic diversity of *Schistosoma haematobium* parasite IS NOT associated with severity of disease in an endemic area in Sudan. BMC Infect. Dis. 14, 469. https://doi.org/10.1186/1471-2334-14-469.

Gatlin, M.R., Black, C.L., Mwinzi, P.N., Secor, W.E., Karanja, D.M., Colley, D.G., 2009. Association of the gene polymorphisms IFN-gamma +874, IL-13 -1055 and IL-4 -590 with patterns of reinfection with *Schistosoma mansoni*. PLoS Negl. Trop. Dis. 3, e375. https://doi.org/10.1371/journal.pntd.0000375.

Grant, A.V., Araujo, M.I., Ponte, E.V., Oliveira, R.R., Gao, P., Cruz, A.A., Barnes, K.C., Beaty, T.H., 2012. Functional polymorphisms in IL13 are protective against high *Schistosoma mansoni* infection intensity in a Brazilian population. PLoS One 7, e35863. https://doi.org/10.1371/journal.pone.0035863.

Gryseels, B., Polman, K., Clerinx, J., Kestens, L., 2006. Human schistosomiasis. Lancet 368, 1106–1118.

Hancock, A.M., Witonsky, D.B., Ehler, E., Alkorta-Aranburu, G., Beall, C., Gebremedhin, A., Sukernik, R., Utermann, G., Pritchard, J., Coop, G., Di Rienzo, A., 2010. Colloquium paper: human adaptations to diet, subsistence, and ecoregion are due to subtle shifts in allele frequency. Proc. Natl. Acad. Sci. U.S.A. 107, 8924–8930. https://doi.org/10.1073/pnas.0914625107.

Hofmann, F., Bernhard, D., Lukowski, R., Weinmeister, P., 2009. cGMP regulated protein kinases (cGK). Handb. Exp. Pharmacol. 137–162. https://doi.org/10.1007/978-3-540-68964-5_8.

Hon, L.S., Zhang, Z., 2007. The roles of binding site arrangement and combinatorial targeting in microRNA repression of gene expression. Genome Biol. 8, R166. https://doi.org/10.1186/gb-2007-8-8-r166.

Hotez, P.J., Kamath, A., 2009. Neglected tropical diseases in sub-saharan Africa: review of their prevalence, distribution, and disease burden. PLoS Negl. Trop. Dis. 3, e412. https://doi.org/10.1371/journal.pntd.0000412.

Isnard, A., Kouriba, B., Doumbo, O., Chevillard, C., 2011. Association of rs7719175, located in the IL13 gene promoter, with *Schistosoma haematobium* infection levels and identification of a susceptibility haplotype. Genes Immun. 12, 31–39. https://doi.org/10.1038/gene.2010.43.

Jackson, J.A., 2015. Immunology in wild nonmodel rodents: An ecological context for studies of health and disease. Parasite Immunol. 37, 220–232. https://doi.org/10.1111/pim.12180.

James, P.M., Coltman, D.W., Murray, B.W., Hamelin, R.C., Sperling, F.A.H., 2011. Spatial genetic structure of a symbiotic beetle-fungal system: toward multi-taxa integrated landscape genetics. PLoS One 6, e25359. https://doi.org/10.1371/journal.pone.0025359.

Jombart, T., 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24, 1403–1405. https://doi.org/10.1093/bioinformatics/btn129.

Jombart, T., Pontier, D., Dufour, A.-B., 2009. Genetic markers in the playground of multivariate analysis. Heredity (Edinb.) 102, 330–341. https://doi.org/10.1038/hdy.2008.130.

Jones, O.R., Wang, J., 2010. COLONY: a program for parentage and sibship inference from multilocus genotype data. Mol. Ecol. Resour. 10, 551–555. https://doi.org/10.1111/j.1755-0998.2009.02787.x.

Joost, S., Vuilleumier, S., Jensen, J.D., Schoville, S., Leempoel, K., Stucki, S., Widmer, I., Melodelima, C., Rolland, J., Manel, S., 2013. Uncovering the genetic basis of adaptive change: On the intersection of landscape genomics and theoretical population genetics. Mol. Ecol. 22, 3659–3665. https://doi.org/10.1111/mec.12352.

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

Kouriba, B., Chevillard, C., Bream, J.H., Argiro, L., Dessein, H., Arnaud, V., Sangare, L., Dabo, A., Beavogui, A.H., Arama, C., Traoré, H.A., Doumbo, O., Dessein, A., 2005. Analysis of the 5q31-q33 locus shows an association between IL13-1055C/T IL-13-591A/G polymorphisms and *Schistosoma haematobium* infections. J. Immunol. 174, 6274–6281.

Leutner, S., Beckmann, S., Grevelding, C.G., 2011. Characterization of the cGMP-dependent protein kinase SmcGK1 of *Schistosoma mansoni*. An. Acad. Bras. Cienc. 83, 637–648.

Leutner, S., Oliveira, K.C., Rotter, B., Beckmann, S., Buro, C., Hahnel, S., Kitajima, J.P., Verjovski-Almeida, S., Winter, P., Grevelding, C.G., 2013. Combinatory microarray and superSAGE analyses identify pairing-dependently transcribed genes in *Schistosoma mansoni* males, including follistatin. PLoS Negl. Trop. Dis. 7. https://doi.org/10.1371/journal.pntd.0002532.

Lotterhos, K.E., Whitlock, M.C., 2015. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. Mol. Ecol. 24, 1031–1046. https://doi.org/10.1111/mec.13100.

Marquet, S., Abel, L., Hillaire, D., Dessein, A., 1999. Full results of the genome-wide scan which localises a locus controlling the intensity of infection by *Schistosoma mansoni* on chromosome 5q31-q33. Eur. J. Hum. Genet. 7, 88–97. https://doi.org/10.1038/sj.ejhg.5200243.

Marquet, S., Abel, L., Hillaire, D., Dessein, H., Kalil, J., Feingold, J., Weissenbach, J., Dessein, A.J., 1996. Genetic localization of a locus controlling the intensity of infection by *Schistosoma mansoni* on chromosome 5q31-q33. Nat. Genet. 14, 181–184. https://doi.org/10.1038/ng1096-181.

Meurs, L., Mbow, M., Vereecken, K., Menten, J., Mboup, S., Polman, K., 2012a. Epidemiology of mixed *Schistosoma mansoni* and *Schistosoma haematobium* infections in northern Senegal. Int. J. Parasitol. 42, 305–311. https://doi.org/10.1016/j.ijpara.2012.02.002.

Meurs, L., Mbow, M., Vereecken, K., Menten, J., Mboup, S., Polman, K., 2012b. Bladder morbidity and hepatic fibrosis in mixed *Schistosoma haematobium* and *S. mansoni* Infections: a population-wide study in Northern Senegal. PLoS Negl. Trop. Dis. 6, e1829. https://doi.org/10.1371/journal.pntd.0001829.

Morel, M., Vanderstraete, M., Hahnel, S., Grevelding, C.G., Dissous, C., 2014. Receptor tyrosine kinases and schistosome reproduction: new targets for chemotherapy. Front. Genet. 5, 238. https://doi.org/10.3389/fgene.2014.00238.

Müller-Myhsok, B., Stelma, F.F., Guissé-Sow, F., Muntau, B., Thye, T., Burchard, G.D., Gryseels, B., Horstmann, R.D., 1997. Further evidence suggesting the presence of a locus, on human chromosome 5q31-q33, influencing the intensity of infection with *Schistosoma mansoni*. Am. J. Hum. Genet. 61, 452–454. https://doi.org/10.1016/S0002-9297(07)64073-7.

Orsini, L., Mergeay, J., Vanoverbeke, J., De Meester, L., 2013. The role of selection in driving landscape genomic structure of the waterflea *Daphnia magna*. Mol. Ecol. 22, 583–601. https://doi.org/10.1111/mec.12117.

Pitchford, R.J., Visser, P.S., 1975. A simple and rapid technique for quantitative estimation of helminth eggs in human and animal excreta with special reference to *Schistosoma* sp. Trans. R. Soc. Trop. Med. Hyg. 69, 318–322.

Polderman, A.M., Mpamila, K., Manshande, J.P., Bouwhuis-Hoogerwerf, M.L., 1985. Methodology and interpretation of parasitological surveillance of intestinal schistosomiasis in Maniema, Kivu Province, Zaire. Ann. Soc. Belg. Med. Trop. (1920) 65, 243–249.

Polman, K., De Vlas, S.J., Van Lieshout, L., Deelder, A.M., Gryseels, B., 2001. Evaluation of density-dependent fecundity in human *Schistosoma mansoni* infections by relating egg counts to circulating antigens through Deming regression. Parasitology 122, 161–167.

R Development Core Team, 2015. R: A language and environment for statistical computing. R Found. Stat. Comput. Vienna Austria. doi:10.1038/sj.hdy.6800737

Rao, C.R., 1964. The use and interpretation of principal component analysis in applied research. Sankhyā Indian J. Stat. Ser. A 26, 329–358.

Rellstab, C., Gugerli, F., Eckert, A.J., Hancock, A.M., Holderegger, R., 2015. A practical guide to environmental association analysis in landscape genomics. Mol. Ecol. 24, 4348–4370. https://doi.org/10.1111/mec.13322.

Restif, O., 2009. Evolutionary epidemiology 20 years on: challenges and prospects. Infect. Genet. Evol. 9, 108–123. https://doi.org/10.1016/j.meegid.2008.09.007.

Richter, J., Hatz, C., Campagne, G., Bergquist, N.R., Enkins, J.M., 1996. Ultrasound in schistosomiasis. A practical guide to the standardized use of ultrasonography for the assessment of schistosomiasis-related morbidity. World Health Organization, Geneva, TDR/STR/SCH/00.1.

Schlossmann, J., Feil, R., Hofmann, F., 2003. Signaling through NO and cGMP-dependent protein kinases. Ann. Med. 35, 21–27.

Selkoe, K.A., Toonen, R.J., 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecol. Lett. 9, 615–629. https://doi.org/10.1111/j.1461-0248.2006.00889.x.

Utzinger, J., Becker, S.L., van Lieshout, L., van Dam, G.J., Knopp, S., 2015. New diagnostic tools in schistosomiasis. Clin. Microbiol. Infect. 21, 529–542. https://doi.org/10.1016/j.cmi.2015.03.014.

Van den Broeck, F., Meurs, L., Raeymaekers, J.A.M., Boon, N., Dieye, T.N., Volckaert, F. A.M., Polman, K., Huyse, T., 2014. Inbreeding within human *Schistosoma mansoni*: do host-specific factors shape the genetic composition of parasite populations? Heredity (Edinb.) 113, 32–41. https://doi.org/10.1038/hdy.2014.13.

Vangestel, C., Mergeay, J., Dawson, D.A., Callens, T., Vandomme, V., Lens, L., 2012. Genetic diversity and population structure in contemporary house sparrow populations along an urbanization gradient. Heredity (Edinb.) 109, 163–172. https://doi.org/10.1038/hdy.2012.26.

Walker, A.J., Ressurreição, M., Rothermel, R., 2014. Exploring the function of protein kinases in schistosomes: perspectives from the laboratory and from comparative genomics. Front. Genet. 5, 229. https://doi.org/10.3389/fgene.2014.00229.

Wilson, S., Jones, F.M., van Dam, G.J., Corstjens, P.L.A.M., Riveau, G., Fitzsimmons, C. M., Sacko, M., Vennervald, B.J., Dunne, D.W., 2014. Human *Schistosoma haematobium* antifecundity ammunity is dependent on transmission intensity and associated with immunoglobulin G1 to worm-derived antigens. J. Infect. Dis. 210, 2009–2016. https://doi.org/10.1093/infdis/jiu374.

World Health Organization, 2006. Preventive chemotherapy in human helminthiasis – coordinated use of anthelminthic drugs in control interventions: a manual for health professionals and programme managers. World Health Organization, Geneva.

Xi, L., Fondufe-Mittendorf, Y., Xia, L., Flatow, J., Widom, J., Wang, J., 2010. Predicting nucleosome positioning using a duration Hidden Markov Model. BMC Bioinformatics 11, 346. https://doi.org/10.1186/1471-2105-11-346.

Zerlotini, A., Aguiar, E.R.G.R., Yu, F., Xu, H., Li, Y., Young, N.D., Gasser, R.B., Protasio, A. V., Berriman, M., Roos, D.S., Kissinger, J.C., Oliveira, G., 2013. SchistoDB: an updated genome resource for the three key schistosomes of humans. Nucleic Acids Res. 41, D728–D731. https://doi.org/10.1093/nar/gks1087.