Check for updates

STUDY PROTOCOL

REVISED Candidate gene family-based and case-control studies of susceptibility to high *Schistosoma mansoni* worm burden in African children: a protocol [version 2; peer review: 2 approved]

Oscar A. Nyangiri[1], Sokouri A. Edwige[2], Mathurin Koffi[2], Estelle Mewamba[3], Gustave Simo[3], Joyce Namulondo[1], Julius Mulindwa[1], Jacent Nassuuna[4], Alison Elliott[4,5], Kévin Karume[6], Dieudonne Mumba[6], P.L.A.M Corstjens[7], M. Casacuberta-Partal[8], G.J. van Dam[8], Bruno Bucheton[9], Harry Noyes[10], Enock Matovu[1], TrypanoGEN+ Research Group of the H3Africa Consortium

[1]College of Veterinary Medicine, Animal Resources and Biosecurity, Makerere University, Kampala, Uganda
[2]Université Jean Lorougnon Guédé (UJLoG) de Daloa, Daloa, Cote d'Ivoire
[3]Faculty of Science, University of Dschang, Dschang, Cameroon
[4]Medical Research Council/Uganda Virus Research Institute and London School of Hygiene & Tropical Medicine Uganda Research Unit, Entebbe, Uganda
[5]London School of Hygiene and Tropical Medicine, London, WC1E, UK
[6]Institut National de Recherche Biomedicale, Kinshasa, Democratic Republic of the Congo
[7]Dept. of Cell and Chemical Biology, Leiden University Medical Center, Leiden, The Netherlands
[8]Dept. of Parasitology, Leiden University Medical Center, Leiden, The Netherlands
[9]Institut de Recherche pour le Développement (IRD), IRD-CIRAD, Montpellier, France
[10]Centre for Genomic Research, University of Liverpool, Liverpool, UK

v2 **First published:** 29 Jun 2021, **4**:36
https://doi.org/10.12688/aasopenres.13203.1
Latest published: 15 Dec 2021, **4**:36
https://doi.org/10.12688/aasopenres.13203.2

Abstract

Background: Approximately 25% of the risk of *Schistosoma mansoni* is associated with host genetic variation. We will test 24 candidate genes, mainly in the T_h2 and T_h17 pathways, for association with *S. mansoni* infection intensity in four African countries, using family based and case-control approaches.
Methods: Children aged 5-15 years will be recruited in *S. mansoni* endemic areas of Ivory Coast, Cameroon, Uganda and the Democratic Republic of Congo (DRC). We will use family based (study 1) and case-control (study 2) designs. Study 1 will take place in Ivory Coast, Cameroon, Uganda and the DRC. We aim to recruit 100 high worm burden families from each country except Uganda, where a previous study recruited at least 40 families. For phenotyping, cases will be defined as the 20% of children in each community with heaviest worm burdens as measured by the circulating cathodic antigen (CCA) assay.

Open Peer Review

Approval Status ✓ ✓

	1	2
version 2 (revision) 15 Dec 2021		✓ view
version 1 29 Jun 2021	✓ view	? view

1. **Severin Donald Kamdem**, University of Utah, Salt Lake City, USA

2. **Samuel K. Armoo**, Council for Scientific

Study 2 will take place in Uganda. We will recruit 500 children in a highly endemic community. For phenotyping, cases will be defined as the 20% of children with heaviest worm burdens as measured by the CAA assay, while controls will be the 20% of infected children with the lightest worm burdens. Deoxyribonucleic acid (DNA) will be genotyped on the Illumina H3Africa SNP (single nucleotide polymorphisms) chip and genotypes will be converted to sets of haplotypes that span the gene region for analysis. We have selected 24 genes for genotyping that are mainly in the Th2 and Th17 pathways and that have variants that have been demonstrated to be or could be associated with *Schistosoma* infection intensity.

Analysis: In the family-based design, we will identify SNP haplotypes disproportionately transmitted to children with high worm burden. Case-control analysis will detect overrepresentation of haplotypes in extreme phenotypes with correction for relatedness by using whole genome principal components.

Keywords

S. mansoni, infection intensity, family based linkage, haplotypes, candidate genes

This article is included in the African Society of Human Genetics gateway.

Corresponding author: Oscar A. Nyangiri (oasanya@gmail.com)

Author roles: Nyangiri OA: Methodology, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Edwige SA**: Methodology, Writing – Review & Editing; **Koffi M**: Conceptualization, Resources, Supervision, Writing – Review & Editing; **Mewamba E**: Methodology, Writing – Review & Editing; **Simo G**: Conceptualization, Resources, Supervision, Writing – Review & Editing; **Namulondo J**: Methodology, Writing – Review & Editing; **Mulindwa J**: Methodology, Writing – Review & Editing; **Nassuuna J**: Methodology, Writing – Review & Editing; **Elliott A**: Conceptualization, Methodology, Resources, Supervision, Writing – Review & Editing; **Karume K**: Methodology, Writing – Review & Editing; **Mumba D**: Conceptualization, Methodology, Resources, Supervision, Writing – Review & Editing; **Corstjens PLAM**: Methodology, Writing – Review & Editing; **Casacuberta-Partal M**: Methodology, Writing – Review & Editing; **van Dam GJ**: Methodology, Writing – Review & Editing; **Bucheton B**: Conceptualization, Methodology, Resources, Supervision, Writing – Review & Editing; **Noyes H**: Conceptualization, Methodology, Supervision, Writing – Review & Editing; **Matovu E**: Conceptualization, Funding Acquisition, Resources, Supervision, Writing – Review & Editing;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported through H3Africa [H3A-18-004]. The Human Heredity and Health in Africa (H3Africa) is a programme of the Alliance for Accelerating Excellence in Science in Africa (AESA) platform. AESA is a funding, agenda-setting and programme management initiative of the African Academy of Sciences (AAS), the African Union Development Agency (AUDA-NEPAD), founding and funding global partners, and through a resolution of the summit of African Union Heads of Governments. H3Africa is supported by the National Institute of Health (NIH), Wellcome, and African Society of Human Genetics (AfSHG). The views expressed herein are those of the author(s) and not necessarily those of AESA and her partners.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2021 Nyangiri OA *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Nyangiri OA, Edwige SA, Koffi M *et al.* **Candidate gene family-based and case-control studies of susceptibility to high *Schistosoma mansoni* worm burden in African children: a protocol [version 2; peer review: 2 approved]** Open Research Africa 2021, **4**:36 https://doi.org/10.12688/aasopenres.13203.2

First published: 29 Jun 2021, **4**:36 https://doi.org/10.12688/aasopenres.13203.1

and Industrial Research, Accra, Ghana

Any reports and responses or comments on the article can be found at the end of the article.

REVISED **Amendments from Version 1**

Following review, we have added additional inclusion criteria. We have revised Figure 1 and added a Supplementary_tableVEP1.xls to describe additional information from SNPs in genes described in Table 1. We further clarify about the diagnostic tests as requested by reviewers in the below response to reviewers.

Any further responses from the reviewers can be found at the end of the article

Abbreviations

BMI: Body mass index

CAA: Circulating anodic antigen

CCA: Circulating cathodic antigen

DRC: Democratic Republic of Congo

FBAT: Family based association test

H3Africa: Human Heredity and Health in Africa research consortium

KK: Kato Katz

LaVIISWA: Lake Victoria Island intervention study on worms and allergy

LOD: Logarithm of odds

MAF: Minor allele frequency

MDA: Mass drug administration

MUAC: Mid Upper arm circumference

PCR: Polymerase chain reaction

POC-CCA: Point of care-circulating cathodic antigen test

PZQ: Praziquantel

QTL : Quantitative trait locus/loci

RDT: Rapid detection test

REC: Research ethics committee

SM1: Schistosoma mansoni susceptibility locus 1

SM2: Schistosoma mansoni susceptibility locus 2

SRR: Sibling Recurrence Risk

TDT: Transmission disequilibrium test

Th1: T helper cell 1 cells/pathway of immunity

Th2: T helper cell 2 cells/pathway of immunity

Th17: T helper 17 cells/pathway of immunity

UCP-LF: Upconverting phosphor lateral flow

UNCST: Uganda National Council of Science and Technology

UVRI: Uganda Virus Research Institute

WHO: World Health Organization

Introduction

Schistosomiasis is a major parasitic disease with prevalence estimates ranging from about 140 million to more than 250 million people worldwide, while nearly one billion are at risk (Colley *et al*., 2014; LoVerde, 2019; McManus *et al*., 2018). In 2016, at least 206.5 million people required preventive treatment and more than 40 million people were treated for the parasite according to the WHO. Currently, 90% of all cases, and most severe cases, are in sub-Saharan Africa (Beltrame *et al*., 2017; Mayaka, 2001). There are 8 to 10 million new cases reported each year worldwide, with 800,000 deaths per year globally and between 200,000 and 400,000 deaths per year in the sub-Saharan African region (Saotoing *et al*., 2011; WHO, 2013).

Schistosomiasis is linked to poverty, affects the socio-economic development of populations (Adenowo *et al*., 2015) and represents a significant burden for developing countries, particularly African countries (Mnkugwe *et al*., 2020). The major species prevalent in African countries are *S. mansoni*, which causes intestinal schistosomiasis and *S. haematobium*, which causes urinary schistosomiasis (Colley *et al*., 2014). School age children have the highest worm burdens (Mnkugwe *et al*., 2020; Saotoing *et al*., 2011).

Mass drug administration (MDA) in school children, snail control programs and improved sanitation are all important to the control of schistosomiasis, but each approach is held back by serious technical difficulties and inadequate resources.

MDA has been the major strategy employed and all children in high prevalence areas are treated irrespective of whether they are infected or not. The frequency of MDA (every six months or every one or two years) is dependent on the prevalence of schistosomiasis in a region (Montresor & World Health Organization, 2011). MDA could be targeted to the most susceptible children; however, there is an incomplete understanding of host genetics of susceptibility, requiring more studies to guide this strategy.

We have reviewed genetic studies of the human response to schistosome infection and found the associations have been reported between schistosome worm burdens and 24 single nucleotide polymorphisms (SNP) in 11 candidate genes (Mewamba *et al*., 2021a). Mapping studies have shown that there are at least three quantitative trait loci (QTL) regions of the genome that regulate the response to infection, however the genes that are responsible for these observations have not been identified (Mewamba *et al*., 2021a). Although candidate gene studies are a powerful strategy for discovering associations between SNP and phenotype, the majority of such observations have not been replicated when studies have been repeated in different

populations (Hirschhorn *et al*., 2002). Whilst two of the 24 SNP associated with schistosomiasis have replicated in additional populations, four have failed to replicate when tested and no attempt has been made to replicate the remaining 18 (Mewamba *et al*., 2021a). We will undertake replication studies for all eleven of these genes in populations from Ivory Coast, Cameroon, Uganda and the Democratic Republic of Congo that are endemic for *S. mansoni* schistosomiasis and that have not been previously included in association studies (Table 1). We will test additional genes that are in QTL and that are known to be involved in the response to schistosome infection but have not been previously included in genetic association studies. We will also test genes in the Th_{17} pathway that have not been included in genetic studies of schistosomiasis. We will genotype using the H3Africa Illumina Omni SNP chip that contains approximately 2.5m SNP. This will enable us to construct haplotypes across whole gene regions for association tests, rather than relying on individual SNP.

Table 1. Genes that will be tested for association with worm burden or egg count. Genes that have been previously associated with schistosomiasis are indicated by references to the relevant publications and those that have not previously been tested in a candidate gene study are marked as Novel. Novel genes are annotated with the number of SNP within 5kb of gene that have had a -log p value > 6 for association with phenotype in any GWAS study, data from GWAS central (27/1/2021) (Beck *et al.,* 2020). A brief justification for their inclusion is shown, a fuller justification can be found in Supplementary Tables and in our review of the genetics of human schistosomiasis (Mewamba *et al.*, 2021a). Other genes that have been considered but are not currently included are *CRP(6), IL9(1), CD14(2), CXCL14 (0), IL3 (0), VEGFA(12), CTGF (0), IL22RA2 (0), NOS3 (4), SHH(1).* The numbers in brackets show the number of SNP with associations in GWAS as in the Novel column in the Table. Amino acid changes functional effects of selected SNP in these genes are in Supplementary_tableVEP1.xls (Noyes, 2021b).

Gene names	Reference (Novel)	Justifications
IFNG	(Gatlin *et al*., 2009)	SM2 Chr6 Region
IL10	(Adedoja *et al*., 2018; Grant *et al*., 2011)	
IL13	(Gatlin *et al*., 2009)(Grant *et al*., 2012; Isnard *et al*., 2011; Kouriba *et al*., 2005)	SM1 Chr5 Region
IL4	(Adedokun *et al*., 2018) (Gatlin *et al*., 2009)	SM1 Chr5 Region
IL5	(Ellis *et al*., 2007)	SM1 Chr5 Region
STAT6	(Adedokun *et al*., 2018; He *et al*., 2008)	
CTLA4	(Idris *et al*., 2012)	
FCN2	(Ouf *et al*., 2012)	
COLEC11	(Antony *et al*., 2015)	
ABO	(Tiongco *et al*., 2018)	
IL12B	Novel (2, Psoriasis)	SM1 chr 5 QTL, Th1 and Th17 pathways
IL17A	Novel (0)	Chr 6 6p21-q21 QTL, implicated in granuloma formation
IL17B	Novel (0)	SM1 chr 5 QTL, implicated in granuloma formation
IL25 (IL17E)	Novel (1)	Implicated in granuloma formation; Th17 pathway
IL17F	Novel (1)	Chr 6 6p21-q21 QTL, implicated in granuloma formation
IL17RA	Novel (5)	Implicated in granuloma formation; Th17 pathway
IL1A	Novel (1)	May regulate Th17 Cell differentiation
IL1B	Novel (6)	May regulate Th17 Cell differentiation
TGFB1	Novel (11)	Regulates Th17 Cell differentiation
IL6R	Novel (20, Asthma, CRP, Fibrinogen)	Chr 1 1p21-q23 QTL Granuloma formation and Th17 pathway
IL6	Novel (9)	May also mediate the Th17 Cell amplification
IL21	Novel (1, Coeliac)	Mediate the Th17 Cell amplification
IL23A	Novel (1, Coeliac)	Stabilize Th17 Cell production
RNASE3 (ECP)	(Eriksson *et al*., 2007)	

The study will sample intensively within a small number of communities in four countries (Ivory Coast, Cameroon, Uganda and the Democratic Republic of Congo) where schistosomiasis prevalence is high. Under these conditions case control studies within communities are at risk of being confounded by cryptic relatedness which can cause type 1 errors (false positives). The large number of candidate gene case control studies that have not been replicated may be partially a consequence of population structure (Mewamba *et al.*, 2021a). We will use two strategies to control for population structure: 1) family-based study designs, which are immune to errors caused by population sub-structure (Laird & Lange, 2009); 2) a case-control design with correction for population structure using principal components calculated from genome wide SNP data.

Estimating intensity of infection is also a problem for schistosomiasis studies. The Kato-Katz method of counting eggs in stool samples is known to lack sensitivity and this is compounded by day to day variation in egg excretions and uneven distribution of eggs across an individual stool (Kongs *et al.*, 2001).

The proposed study aims to improve on previous studies by using:

1) A more reproducible phenotyping method (CCA or CAA)
2) A more systematic selection of candidate genes (Table 1)
3) A study design that is less sensitive to population structure

The specific study aims are as follows:

1) Test candidate genes listed in Table 1 for association with intensity of *S. mansoni* infection
2) Test if previously published associations are replicated in other populations

The hypotheses we will test are:

i) Variation in Th2 and Th17 pathway genes contribute to variation in *S. mansoni* worm burden
ii) Genes in the Th17 pathway contribute to the association with schistosomiasis at three QTL
iii) Associations found with genes in the SM1 and SM2 QTL are replicable across multiple populations

Protocol

Ethical approval and consent to participate

The study will be carried out in Cameroon, Uganda, DRC and Ivory Coast. In Cameroon the study was approved by the National Ethics Committee for research on human health of the Ministry of Public Health of Cameroon with the reference number N°2019/02/1144/CE/CNERSH/SP. The review board of the Molecular Parasitology and Entomology Subunit of the Department of Biochemistry of the Faculty of Science of the University of Dschang gave its approval.

In Uganda, the approval was from the Uganda National Council for Science and Technology (UNCST) number: HS 1183 and the Uganda Virus Research Institute Research Ethics Committee: approval of amendment to perform genetic study on human susceptibility to schistosomiasis Ref: GC/127/18/09/317

In DRC, the ethics review number is COMETH/TKK/PRES/004/2018 approved by the Clinic Ngaliema Ethical Committee of the Health Ministry.

In Ivory Coast, the study was approved by the National Ethics Committee for Life Sciences and Health (cnesvs); approval number N/Ref: 040-19/MSHP/CNESVS-kp. In all countries, since schools are involved, we will consult the ministry of education in conducting the study.

Study design

Two study designs will be employed in different countries, described in detail below.

Study 1: Family based linkage study; to be conducted in the Democratic Republic of Congo, Ivory Coast, Cameroon and the Uganda Lake Victoria islands.

Study 2: Case control study; to be conducted in the Lake Albert Region of Uganda.

Study 1: Family based linkage and association study

Study sites

Uganda: previous collections by the Lake Victoria Island Intervention Study on Worms and Allergy (LaVIISWA), conducted on the Lake Victoria islands of Koome.

Ivory Coast: Blolquin and Guiglo (West)

Cameroon: Makenene and Adamawa regions

DRC: Kongo Central (Kisantu and Kimpese)

Participant recruitment

Consent will be obtained in two stages.

1) Written collective consent for screening for schistosome worm burden will be provided by school directors or health authorities.
2) Written informed consent for children who meet inclusion criteria to participate in the study will be requested from parents or guardians in the home after children have been screened for intensity of infection at school.

The informed consent forms and case report forms are provided as part of extended data in Harvard Dataverse (Noyes, 2021a).

Community entry procedures differ in each country as mandated by the respective ethical review committees. In DRC administrative authorities, school inspectors, school directors and

teachers were involved, whereas in Ivory Coast administrative authorities, Regional Health Director, Departmental Health Director school inspectors, school directors and teachers were involved. In Cameroon, the following were notified about the study; Ministry of Health and Ministry for Schools for Region, school inspectors, school directors and teachers in participating schools. In Uganda, the ministry of health and village health teams were involved in community entry.

Subsequently collective consent will be obtained from district health authorities or school directors and the date of the screening will be planned. The parents or guardians of children at schools will be informed by school directors of schistosomiasis screening at the school on a given date and invited to let their children participate. In Cameroon, DRC and Ivory Coast, schools were chosen that were in high *S. mansoni* endemic areas. The number of schools is not defined at the beginning as it is dependent on the number of high worm burden families that can be recruited in the school. We will continue to sample schools until we have recruited our target number of children, which is at least 100 families per country.

After screening children for intensity of infection, parents or guardians of children who meet inclusion criteria will be visited and invited to provide informed consent for themselves and their participating children. Children older than 10 years will be invited to provide assent. Each participant will be free to choose whether to join the study after the objectives, procedures, and potential risks and benefits of the study have been clearly explained to them using lay terms.

Screening procedures. Schistosomiasis screening will be performed at schools by a team from TrypanoGEN+ together with the National Control Programs team. Children who agree to participate in the study will be tested for *S. mansoni* infection by detecting circulating cathodic antigen in their urine and *Schistosoma mansoni* eggs in their stools, using the Point Of Care Circulating Cathodic Antigen (POC-CCA) test (van Dam *et al.*, 2004) and the Kato Katz (KK) technique respectively (World Health Organisation, 1994). Urine and stool will be collected in plastic collection tubes and bags. In each country, some samples may be kept in the institution which will conduct the research, i.e. University of Dschang (Cameroon samples), INRB (DRC) and University Jean Lorougnon Guédé University (Ivory Coast). All children will be treated with Praziquantel by the national Ministry of Health Mass Drug Administration teams according to the recommendation of the World Health Organization (World Health Organisation, 2006).

Inclusion criteria

1. Age between 5–15 years
2. Intensity of infection in the top 20% of positive children in their community
3. Have at least one parent available for inclusion in the study
4. Born in the community or spent at least a year in the community

Exclusion criteria

1. No parent available
2. Children positive for *S. haematobium* by urine filtration test or PCR
3. Treatment with praziquantel for schistosomiasis in the last six months

Procedures for the top 20% index children in families. The families of the children bearing high worm burden infections will be visited in their homes to meet their parent or guardian and inform them about the objectives and goals of the study. The head of the families that meet the inclusion and not the exclusion criteria will be invited to give their consent for their family to participate to the study. Adult participants will also be invited to give their consent and the index child and siblings will give their assent if over 10 years. Children, who are encountered in the home and have not been screened at school and meet the enrolment criteria, will be tested for inclusion in the study. Each person who gives consent or assent and meets the enrolment criteria will be asked to provide about 5ml of blood. This will be done in the home and samples will be kept at cold storage for transportation to the relevant laboratory; University of Dschang (Cameroon samples), INRB (DRC) and University Jean Lorougnon Guédé University (Ivory Coast). Figure 1 shows a flow-chart of the procedures for this study.

Timelines. Timelines for the study are shown in Table 2. Initially, all work was projected to take 3 years. However, collections started in 2019 and may continue until 2023 due to restrictions imposed by the coronavirus disease 2019 (COVID-19) pandemic.

Diagnosis

Kato Katz test

Each participant will provide a stool sample for the Kato-Katz test (KK). The WHO approved standard KK test will be conducted (Katz *et al.*, 1972). To assure of the quality of the test, we will homogenize stool by mixing in a polythene bag to make the distribution more uniform. After this we will pick the stool using spatula from different areas of the stool sample and make the Kato Katz stool slide. We will take duplicate slides and have two separate readings of egg counts by technicians. The average of these readings is then taken as the eggs per gram.

This technique can be used to distinguish the different helminth species eggs (*S. mansoni Ascaris lumbricoides*, *Trichuris trichiura*, hookworm and *Taenia* spp), which will be recorded separately.

Circulating cathodic antigen (CCA) test

The KK thick stool smear is the most common field technique for *S. mansoni* identification and quantification. This method

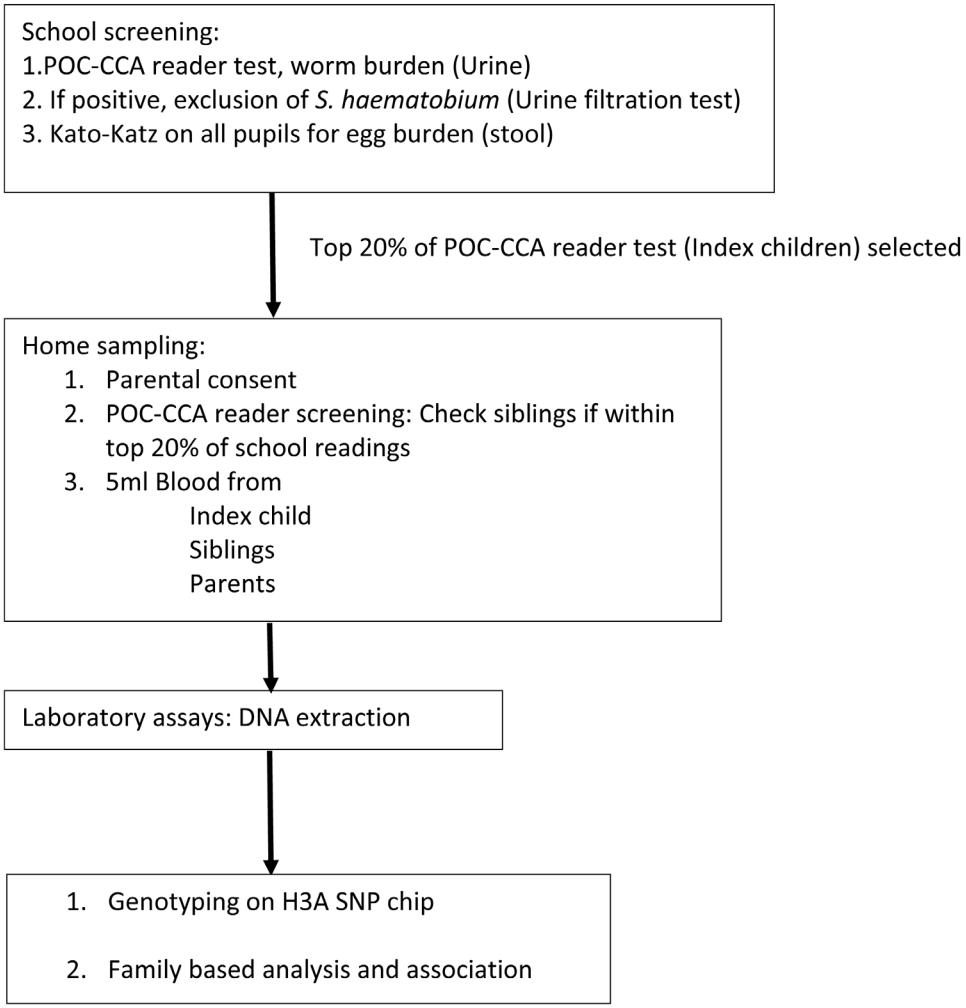

Figure 1. Study flow chart for family-based association study (FBAT) studies. POC-CCA=point of care-circulating cathodic antigen; PCR=polymerase chain reaction; DNA=deoxyribonucleic acid; SNP=single nucleotide polymorphisms.

Table 2. Timeline for the study.

Event	Year1	Year 2	Year 3
Ethical clearance			
Administrative approval			
Sample collection			
Family/case-control inclusion			
DNA extraction			
Genotyping			
Data analysis			
Publication of manuscripts			

is 100% specific, however it has poor sensitivity for detecting low intensity infections (Colley *et al.*, 2017). More recently the CCA that is produced by adult worms (Nash & Deelder, 1985) and is detectable in blood and urine, has been used as the basis for a lateral flow test that can be used at the point of care (Kittur *et al.*, 2016; WHO, 2012; van Dam *et al.*, 2004). The advantages of the POC-CCA test include its rapid procedure, easy handling under field conditions and minimal practical training requirements for its application (Glinz *et al.*, 2010). Several studies have shown that this test appears to be more sensitive than the KK technique (Deelder *et al.*, 1976; Kittur *et al.*, 2016; Mnkugwe *et al.*, 2020). Nevertheless, the specificity of POC-CCA is influenced by cross-reactivity with other helminth infections and, more importantly, may produce false positives with weak bands especially in areas of moderate and low prevalence and this can lead to overestimation of prevalence

(Colley *et al.*, 2017). However in areas of high transmission, such as those where studies 1 and 2 are being undertaken, all trace positives should be treated as positives (Colley *et al.*, 2020). Although the abundance of CCA antigen is believed to be proportional to the number of worms present (van Dam *et al.*, 2004), the POC-CCA is marketed as a qualitative test for *S. mansoni* diagnosis and not for worm quantification. However, three studies have demonstrated that there is a semi-quantitative relationship between band intensity estimated visually and egg count on the Kato-Katz test (Coulibaly *et al.*, 2013; Dawson *et al.*, 2013; Kittur *et al.*, 2016). This has recently been improved with a set of 10 standards of reference cassettes (G-scores) with ink jet printed bands of increasing intensity (Casacuberta-Partal *et al.*, 2019). We have recently extended this method by using the ESEQuant LR3reader to read the intensity of the positive band. In field trials ESEQuant readings correlated well with estimates of band intensity using the reference cassettes developed by Casacuberta-Partal *et al.* (2019) with a Spearman rank correlation coefficient = 0.89 (Mewamba *et al.*, 2021). We will therefore use the ESEQuant LR3 reader to objectively quantify band intensity on POC-CCA cassettes in a standardized way at all sites in Study 1 except the samples from Lake Victoria. The samples from the Uganda Lake Victoria site are already archived and worm burdens will be quantified using the more sensitive CAA assay.

Using the RDT-Reader for quantitative POC-CCA

The ESEQuant LR3 Gold reader (part number ESLR11-MB-6401 with a drawer inlay for the POC-CCA cassette part number ESLR05-MA-5167 DIALUNOX GmbH, Germany) will be used to measure reflective signals on POC-CCA cassettes. This gives an objective measure different from the subjective visual results. First, a method is developed on the reader to capture valid signals for the control line and test lines. This method will be deployed on subsequent RDT tests on this reader. A quantitative test will be done on the POC-CCA strip by spotting a drop of urine on the POC-CCA. The sample number will be pre-entered on the reader, and once 20 minutes elapse the cassette will be put into the reader and the method will be initiated to read the cassette. Cassettes that are not read at 20 minutes will be discarded and the test will be repeated. An OD signal indicating the intensity of signal will be shown. The intensity per sample is stored on the reader and can be downloaded to compare signal intensities across various samples.

Circulating anodic antigen (UCP-LF test)

The upconverted phosphor lateral flow (UCP-LF) is a highly sensitive assay that is used to detect the CAA antigen regurgitated by adult *Schistosoma* worms and is present in urine, plasma and serum (van Dam *et al.*, 2013). The CAA is much more abundant in plasma than CCA which in combination with its high specificity and fluorescence reading technology leads to a test with much higher sensitivity.

We will use the schistosome CAA assay that requires 20µl of plasma (SCAA20). Since the POC-CCA assay has only been used as a semi-quantitative test (Coulibaly *et al.*, 2013; Dawson *et al.*, 2013; Kittur *et al.*, 2016) we will validate the CCA test against the CAA test in 100 samples. A previous comparison of the two methods in a low endemicity community in Brazil found only modest correlation (Spearman's rho = 0.24) between CAA and CCA readings which were classified as negative, trace or positive (Sousa *et al.*, 2019). In that study 36% of samples were positive by CAA, 12% by CCA and 1.6% by Kato-Katz demonstrating the relative sensitivity of the methods. CAA results will be compared against 100 CCA results for study 1 to validate CCA reader signals ability to pick out the top 20% of worm burden. We calculated the power to detect a correlation between CAA and CCA signal intensities using the pwr.r.test function in the R pwr library. 100 samples will have 80% power to detect a correlation (r> 0.28). We will also use the CAA test as the primary measure of worm burden for the Lake Victoria samples.

Human negative plasma will be spiked with a known concentration of CAA and dilutions made. A negative plasma sample that has not been spiked with CAA will be included as a negative control. Plasma samples and standards will be extracted with an equal volume of 4% w/v trichloroacetic acid, vortexed and incubated for 5 minutes. Of the resulting TCA soluble fraction, 20µL will be added to wells coated with UCP particles and anti-mouse monoclonal anti-CAA antibodies hydrated with 100ul of high salt lateral flow buffer. The wells will then be incubated for one hour at 37°C while shaking at 900rpm in a shaking incubator. After one hour, the lateral flow strips will be placed in UCP bound wells and the samples will be allowed to flow. The strips will be dried overnight and analysed using the Labrox Upcon scanner (serial number 2180023). The test line (T) signals will be normalized to the flow control (FC) signals of the individual strips and the results expressed as ratio values i.e. T/FC.

Urine filtration and PCR to exclude *S. haematobium*

Since *S. mansoni* CCA can also cross react with *S. haematobium* CCA, all CCA positive samples should be processed further to exclude *S. haematobium* co-infected samples. This will leave only the *S. mansoni* mono-infected samples for analysis. Urine filtration followed by microscopy will be used in the field as a screening test to exclude *S. haematobium* co-infected samples. All *S. haematobium* negative samples in the urine filtration will be further tested with *S. haematobium* specific PCR, a more sensitive test to exclude *S. haematobium* co-infection. Where urine filtration test is not available, all POC-CCA positive urine samples will directly undergo the PCR test without prior screening with the urine filtration test. The *S. haematobium* PCR works best on urine concentrates. *S. haematobium* positive samples will be identified by amplifying the repetitive *Dra1* sequence which is specific to S. *haematobium* using primers identified by Hamburger *et al.*, 2001 (Hamburger *et al.*, 2001).

Genotyping and association testing

Candidate genes

Whilst many genes are known to participate in the response to schistosome infection (Kamdem *et al.*, 2018) there is a more limited set of genes that may have variants that contribute

to the outcome of infection (Mewamba *et al*., 2021a). A list of genes that will be tested is shown in Table 1. Candidate genes have been selected on three principles: 1) genes that have previously been shown to be associated with schistosomiasis in other studies; 2) genes in the QTL regions that have been reported to be involved in the response to schistosome infection; 3) genes that are involved in the Th_{17} pathway. We have previously shown that Th_{17} pathway genes are present in multiple quantitative trait loci and may contribute to the effect associated with these loci (Mewamba *et al*., 2021a). Further justification and functional effects of selected SNPs for the genes selected is in the extended data (Noyes, 2021b) and in (Mewamba *et al*., 2021a).

Genotyping

DNA will be extracted from whole blood from index cases and their family members using the Qiagen mini kit (QIAamp DNA minikit catalogue number 51306, Qiagen, Germany) or the Quick-DNATM Miniprep Plus Kit (Catalogue Number: D4069, Zymo research, USA) according to manufacturer's instructions. Genotyping will be done by Illumina (San Diego) on the Illumina H3Africa Pan African V2 chip (The H3Africa Consortium *et al*., 2014) which contains approximately 2.5 million SNPs enriched for SNP that are informative in African populations.

Family based analyses

We will use family based analysis methods as reviewed by Ott *et al.* (Ott *et al*., 2011). The method applied will depend on the structure of families collected. Here we describe one such method, the transmission disequilibrium test (TDT). The TDT assesses evidence of association between markers and schistosomiasis in the presence of linkage. The test statistic (also known as McNemar's Test) is chi squared distributed $\chi^2 = \frac{(T-NT)^2}{(T+NT)}$ where T is the number of times the allele of interest is transmitted from heterozygous parents to affected offspring and NT is the number of times that allele is not transmitted (Spielman *et al*., 1993). We will use the family based association test (FBAT V1.4) implementation of the TDT (Laird *et al*., 2000) to test each haplotype allele in each haplotype block for association with schistosomiasis. If there are sufficient families with more than one affected sib, then we will also use the affected sib-pair test in the same package. We will use the multi-locus transmission disequilibrium test (M-TDT) method (Loucoubar *et al*., 2017) to search for evidence of additive and epistatic effects of multiple loci.

Meta-analysis

Initially data sets from each country with at least 100 families will be analysed independently to identify loci that are associated with populations in each country. Subsequently the data from all four participating countries will be combined in order to undertake a meta-analysis and identify loci that are of smaller effect than are detectable in local studies and that are also consistently associated with disease in multiple African countries. We will use R version 4 for plotting graphics and supplementary data, FBAT V1.4 will be used for fbat analysis and Plink V1.90 for transmission disequilibrium tests (TDT).

Association analysis with haplotypes

Family based methods for the discovery of linkage and association depend on being able to calculate the probability of the child genotypes given the genotypes of the parents. These calculations are only meaningful if there is more than one allele that the children might inherit from each parent, ie the parents need to be heterozygous in order for them to be informative at a particular locus.

Since SNP usually only have two alleles the frequencies of heterozygote parents will generally be low. For example, with the highest possible minor allele frequency (MAF) of 50%, 50% of parents will be heterozygous and only 25% of families will have two heterozygous parents. The frequency of heterozygotes can be increased by using SNP haplotypes that also have more alleles. In order to test the viability of using haplotypes, rather than SNP, for the association analysis we ran a simulation using 1,000 Genomes Project human sequence data (Abecasis *et al*., 2012) to discover the empirical distribution of haplotype frequencies. Having identified 378 haplotype blocks within the 24 candidate genes we ran a power analysis to discover the proportion of loci where there was adequate power to discover different values of haplotype relative risk (supplementary text in extended data (Noyes, 2021b).

The power analysis found the number of families that would be required to detect haplotype relative risks (HRR) of 1.5, 2 and 3 with haplotype allele frequencies from 0.05 to 0.95 after correction for the 378 haplotype blocks being tested (extended data Figure 2A (Noyes, 2021b). The analysis showed that 250 families would provide at least 80% power to detect a HRR of 2 for haplotypes with allele frequencies between 16% and 72%. At 80% of haplotype blocks over 50% of chromosomes were on a haplotype with a frequency in the informative range (extended data Figure 2B (Noyes, 2021b). However, at 46 out of the 378 loci no haplotypes had allele frequencies in this range (extended data Figure 2C (Noyes, 2021b) and therefore only large HRR (HRR > 2) are likely to be detectable at these loci. In summary 250 families will provide at least 80% power to detect HRR > 2 at the 60% of blocks that have haplotypes with allele frequencies between 16 and 72%. The number of alleles and allele frequencies in a haplotype block can be controlled by changing the block length. We used a standard block length of four SNP loci for our simulation but the optimum number for maximum potential power will vary between loci depending on allele frequencies and local linkage disequilibrium. Prior to running the association tests, we will use the genotype data to optimise the sets of haplotype blocks within each gene to maximise theoretical power to detect associations.

Consent to publish. All authors will approve the final manuscript before publication. No institutional consent is required for publication.

Data management. Study documents will be kept securely in the institution which will conduct the research, i.e. University of Dschang (Cameroon samples), INRB (DRC), Makerere University(Uganda samples) and University Jean Lorougnon Guédé University (Ivory Coast). These will be in physical

lockable cabinets for forms and password protected databases with only study staff and investigators having access. All data will be anonymized before publication in appropriate scientific journals. Accompanying genotype data will also be anonymized and will be deposited in an open access depository.

Data will be collected in standardised forms as shown in supplementary data. It will be entered in Excel sheets and SQL database with range checks. The variables in the case report forms will be precoded for categorical variables to allow proper analysis. Physical data forms will be stored securely. Only the study investigators and project staff involved in generating the data will have access to the data.

Dissemination of study outcomes. No actionable information is expected from the association analysis and individual results will not be returned to participants. General feedback will be shared with study participants and community leaders at public meetings and results shared with policy stakeholders including officials of the National Schistosomiasis Control Programs and Ministries of Health. Sponsors of the research will receive periodic reports on study progress. Analyses will be performed separately for each study population and published independently so that national authorities can have a clear understanding of the genetic risk factors for schistosomiasis in their countries. A meta-analysis of the combined data will also be undertaken and published separately.

Study 2: Case-control association study

Study site: Lake Albert region, Uganda. The major difference between Study 1 and Study 2 is the study design. They therefore share the schistosomiasis specific methods described above for study 1; the CCA test, Kato Katz, DNA extraction and genotyping and the panel of genes to be tested. Study 2 will differ from study 1 in recruitment procedures including inclusion criteria, definition of cases and controls and data analysis methods. Below we describe what will be the different in study 2.

Recruitment and case definition. In the study site, the population will be screened using the CCA rapid test with results scored visually as trace, +1, +2, +3 (see section on CCA above) (Casacuberta-Partal *et al*., 2019). All individuals will be invited to give a blood sample for DNA extraction and plasma for subsequent worm burden quantification by CAA.

Extreme phenotype sampling will be done, where only the genotypes of the top 25% and bottom 25% of worm burden will be analyzed as described by Li *et al.* (Li *et al*., 2019). Genotyping only the participants with extreme phenotypes reduces the genotyping cost by 50% but has very little effect on power (Li *et al*., 2019). To allow determination of these extreme phenotypes, worm burden will be determined on plasma using the CAA test in the laboratory at UVRI. The CAA test has been described above. Since the case-control design requires controls to be unrelated to the cases, the 25% of participants in the top and bottom tails of the CAA distributions will be included for genotyping. The samples will be genotyped using the H3A chip defined above. The genotype data will be used to filter out individuals who are first cousins or closer and share a phenotype to minimize bias.

After filtering, the target proportions of the total sample for cases and controls is expected to be at least 20% of the original population sampled. In a multiplicative model; a 20% fraction from each extreme phenotype would give 80% power to detect 35% increase in risk; for MAF 0.3 with sample size 300 (60 cases and 60 controls genotyped) (Li *et al*., 2019). Most effects in the classical SM1 locus QTL have been shown to be co-dominant (Abel *et al*., 1991; Marquet *et al*., 1996) and therefore a multiplicative model is appropriate.

We will therefore aim to recruit at least 500 CAA positive children in Uganda. The case-control recruitment profile is shown in Figure 2

Additional data recorded per participant. The following features will be recorded at the time of sample collection as potential covariates during analysis: age, sex, fever, diarrhea, infections with other parasitic worms will be detected by microscopy as part of the KK procedure, clinical observation for distended abdomen, Mid Upper Arm Circumference (MUAC) and body mass index,BMI (weight divided by the square of height in metres). The case report is provided as part of the extended data (Noyes, 2021a).

Data analysis. A candidate gene case control linear regression analysis will be undertaken in Plink v1.9 and R 4.0, with haplotypes as the independent variable and CAA as the dependent variable. Results will be controlled for population structure using principal components also calculated in Plink v1.9 as covariates. In this sampling design, it is critical to control for population structure (Panarella & Burkett, 2019), and genotyping on the whole genome H3A chip will provide sufficient data for a calculation of principal components for this purpose. Age, sex, MUAC and BMI will also be used as additional covariates. Procedures for the KK test will be the same as those of study 1.

Dissemination of study outcomes. Study participants will immediately receive their results for testing of schistosomiasis and will be advised on control measures by the health team. Since the national control programs are also involved in the study, they will have access to the results of the study for use in programmatic operations and policy decisions.

Selection of candidate genes to be tested

The set of candidate genes that we will test (Table 1) is based on our recent review of all previous studies of genetic variants associated with schistosomiasis infection intensity (Mewamba *et al*., 2021a). We included candidate genes based on three sets of overlapping criteria:

1) Eleven genes that have been found associated with schistosomiasis in other populations (*IFNG, IL10, IL13,*

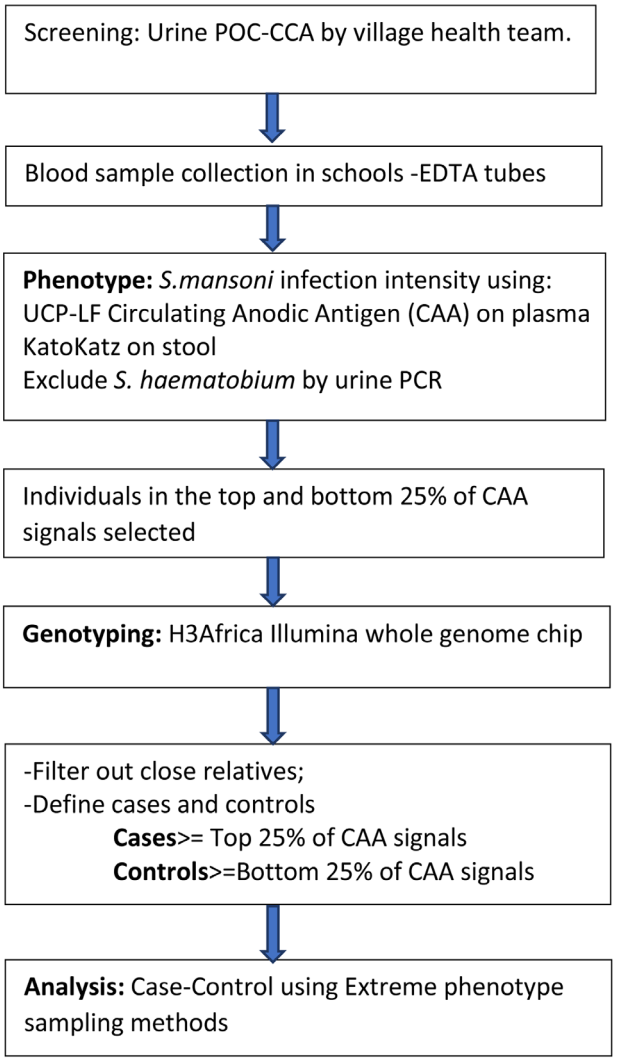

Figure 2. Study profile for the case-control study. POC-CCA=point of care-circulating cathodic antigen.

IL4, IL5, STAT6, CTLA4, FCN2, COLEC11, ABO, RNASE3)

2) Five genes in schistosomiasis quantitative trait loci (QTL) that are known to participate in the response to schistosomiasis infection but have not previously been tested in candidate gene studies (*IL17A, IL17B, IL17F, IL6R, IL12B*)

3) Eight genes in the Th_{17} pathway (*IL1A, IL1B, TGFB1, IL6, IL21, IL23A, IL25, IL17RA*)

SNP in only four of the eleven previously tested genes in Table 1 have been found associated with infection with schistosomes in more than one population, these genes are indicated by multiple citations in Table 1. It is important to replicate the findings for the other genes since the earlier studies were generally small and candidate gene studies are known to be poorly reproducible (Hirschhorn *et al.*, 2002). Amino acid changes and functional effects of selected SNP in these genes are in Supplementary_tableVEP1.tab (Noyes, 2021b).

Four quantitative trait loci (QTL) associated with *S. mansoni* egg count have been discovered (Marquet *et al.*, 1996; Marquet *et al.*, 1999; Müller-Myhsok *et al.*, 1997; Zinn-Justin *et al.*, 2001). These QTL were discovered before the human genome provided lists of genes in these regions. We systematically scanned these regions for genes that have been studied in the context of schistosomiasis (Mewamba *et al.*, 2021a). Genes in QTL with strong evidence for involvement in the response to infection have been included in this study. These genes are annotated with the QTL location in Table 1.

Three of the four QTL regions contained genes in the Th_{17} pathway. The Th_{17} system has two main functions: it regulates

the clearance of extracellular pathogens, it also helps the B cells to induce tissue inflammation (Patel & Kuchroo, 2015). Although the Th_{17} system is known to be involved in the response to schistosomiasis (Larkin *et al.*, 2012; Mbow *et al.*, 2013), no studies of the genetics of Th_{17} pathway genes in human schistosomiasis have been conducted so far. We therefore included Th_{17} genes in the QTL regions in our study (Table 1) as well as important Th_{17} genes that were outside the QTL (*IL1A, IL1B*, *IL17RA, IL21, IL23, IL25, IL6, TGFB1*).

By testing for association with schistosomiasis in four different populations we will be able to identify genes which have variants which contribute to the heaviest worm burdens across Africa. Other genes may be of local importance or previously observed associations might be false positives as a consequence of sampling from stratified populations. Genetic data has recently been used to compile risk scores for developing schistosomiasis related sever liver disease (Dessein *et al.*, 2020). With the data from this study, it may be possible to develop a similar risk score for acquiring high worm burdens.

Discussion

There is abundant evidence for genetic contributions to the risk of acquiring schistosomiasis and to having particularly heavy worm burdens (Dessein *et al.*, 2020; Mewamba *et al.*, 2021a). Some of this evidence has come from candidate gene studies that tested for association with genes that were known to be involved in the response to infection, but these studies have rarely been replicated and, where they have, the results have often been inconsistent. Other evidence has come from mapping studies that have identified QTL, regions of the genome that contain one or more genes with variants that change the risk of schistosomiasis. However, in most cases, no genes within these regions have been tested for association with disease. We have systematically surveyed these genes for plausible candidate genes which are known to participate in the response to infection (Mewamba *et al.*, 2021a) and observed that Th17 pathway genes were found within the majority of these regions.

We will genotype the samples on the H3Africa Whole Genome SNP chip making it possible to scan for associations with any gene in the genome in a Genome Wide Association Study (GWAS). However large numbers of samples are required for a GWAS to compensate for the large number of tests being conducted and, since we only expect to collect 100 families per population for family-based analyses and 100 cases and 100 controls for the case control study, we would only have the power to detect exceptionally large relative risks in a GWAS. By declaring the list of candidate genes in advance of analyzing the data we can establish a much more limited set of hypotheses which we will have the power to test with the number of samples available.

Study status

The study is still recruiting participants. The current progress is: Cameroon (100 families consisting of 411 individuals); Ivory Coast (10 families consisting of 39 individuals); Democratic Republic of Congo (14 families consisting of 40 individuals); Uganda (LaVIISWA; previous family based collection, 40 families; Albertine case control collection: 600 individuals).

Data availability

Underlying data

No data are associated with this article.

Extended data

Harvard Dataverse: Consent forms and CRFs for family based or case control schisto studies. https://doi.org/10.7910/DVN/6OLLGW (Noyes, 2021a).

This project contains the following extended data:

- CAM_autorisation recherche schisto.pdf (Cameroon health ministry authorisation for research)
- CAM_Clairence éthique schistosomiase.pdf (Cameroon ethics review clearance)
- CAM_Consent form_Anglais.pdf (Cameroon consent form in English)
- CAM_Consent form_french.pdf (Cameroon Consent form in French)
- CAM_Information sheet.pdf (Cameroon information sheet)
- CIV_Assent form.docx (Ivory Coast Assent form English)
- CIV_Consentements_Projet TrypanoGen+.docx (Ivory Coast information sheet, consent forms in French)
- CIV_Informed Consent Form.docx (Ivory Coast Consent form in English)
- DRC_collective_consentform_and_CRFs.docx (Democratic Republic of Congo Consent form and case report form)
- DRC_Fiche de Consentement et Information.docx (DRC information sheet)
- DRC_Fiche de Consentement Personnel.docx (DRC personal consent form)
- UGA_LaVIISWA_consent_form.pdf (Uganda LaVIISWA study consent form: Refer for future use allowing genetic studies)
- UGA_SCHISTO Assent.pdf (Uganda assent form for the case-control study)
- UGA_Schisto Consent.pdf (Uganda case control study consent form)
- UGA_SCHISTO CRF.pdf (Uganda case control study case report form)
- UGA_Schisto Sample Storage.pdf (Uganda case control study sample storage consent)

Harvard Dataverse: Haplotype Genotyping. https://doi.org/10.7910/DVN/OVQPCO (Noyes, 2021b).

This project contains the following extended data:

- Amino acid changes and biological consequences associated with SNP selected from genes in Table 1
  - Supplementary_tableVEP1.xlsx Selected SNP functional effects from Variant effect predictor tool http://grch37.ensembl.org/info/docs/tools/vep/
- Files for modelling the effect of using haplotypes for transmission disequilibrium test
  - SupplementaryMaterial_Association_analysis_with_haplotypes.pdf (Description of association analysis with haplotypes and the power to detect different haplotype relative risks given different haplotype allele frequencies)
  - SupplementaryTables_1KG_African_haplotype_blocks.xlsx (Excel file containing lists of haplotype blocks identified in candidate genes in One Thousand genomes African samples, together with the number of samples required to have 80% power to detect Haplotype relative Risk of 1.5 or 2.0 or 3.0. Also, a worksheet with details of each haplotype in each block)
- Software
  - ReadMe.txt (Description of how to run the pipeline)
  - makeHaplotypes.pl (perl script called by runBigLD.sh to obtain the SNP in each haplotype block, identify all the haplotypes in a block and count their frequencies. It outputs Summary.HapStats.1kg.h3a.POP.txt containing summary statistics on each haplotype block in each gene in a population and All.HapsStats.1kg.h3a.POP.txt containing the sequence of each haplotype in each haplotype block in each candidate gene and counts of observations of each haplotype. POP is a population name)
  - plotBlocks.R (Rscript to make plots of haplotype blocks for each 1000 genomes African population)
  - replaceFamIds.pl (Perl script to modify 1000 genome sample ids so that they ids include a three-letter population identifier. Uses ids and population names in OneKg.pops. Takes a fam file with 1000 genome ids as input parameter)
  - runBigLD.R (Rscript that uses the BigLD R package to identify haplotype blocks and output a linkage disequilibrium heatmap and a text file of haplotype block co-ordinates. The text file is subsequently deleted)
  - runBigLD.sh (Wrapper shell script that prepares input files for runBigLD.R, calls runBigLD.R and also calls makeHaplotypes.pl to process runBigLD.R output)
- Data files used in input
  - OneKg.pops (Text file with all 1000 genome ids and the populations to which they belong. Used by replaceFamids.pl)
  - exons.ensGRCh37.txt (Co-ordinates of exons for use by plotBlocks.R)
  - getExons.sh (shell script to extract exons for just genes of interest from a larger mart_export.txt (not provided) file which should be downloaded from Biomart. Note the script makes assumptions about the columns in which data will be found and these should be adjusted according to the mart_export.txt file used)
  - SchistoCandGenes.1kg.h3a.bed (Plink bed file with genotypes of SNP that are present in both the 1,000 genomes dataset and the Illumina H3Africa genotyping chip and that are in the candidate genes analysed)
  - SchistoCandGenes.1kg.h3a.bim (Plink bim file with identifiers of SNP that are present in both the 1,000 genomes dataset and the Illumina H3Africa genotyping chip and that are in the candidate genes analysed)
  - SchistoCandGenes.1kg.h3a.fam (Plink fam file with identifiers for 1000 genomes African samples)
- Output from simulations for each country and combined
  - All.HapsStats.1kg.h3a.ALL.txt (Sequence of each haplotype in each haplotype block in each candidate gene and counts of observations of each haplotype for all populations (GWD, MSL, YRI, ESN, LWK))
  - All.HapsStats.1kg.h3a.ESN.txt (Sequence of each haplotype in each haplotype block in each candidate gene and counts of observations of each haplotype for ESN population)
  - All.HapsStats.1kg.h3a.GWD.txt (Sequence of each haplotype in each haplotype block in each candidate gene and counts of observations of each haplotype for GWD population)
  - All.HapsStats.1kg.h3a.LWK.txt (Sequence of each haplotype in each haplotype block in each candidate gene and counts of observations of each haplotype for LWK population)
  - All.HapsStats.1kg.h3a.MSL.txt (Sequence of each haplotype in each haplotype block in each candidate gene and counts of observations of each haplotype for MSL population)
  - All.HapsStats.1kg.h3a.YRI.txt (Sequence of each haplotype in each haplotype block in each

candidate gene and counts of observations of each haplotype for YRI population)

- GeneHapBlockPlots.pdf (Output from plotBlocks.R)
- Heatmaps.tif.tar.gz (Heatmaps showing linkage disequilibrium and haplotype blocks within each gene. The heatmaps are output from the runBigLD.R Rscript)
- Summary.HapStats.1kg.h3a.ALL.txt (Summary statistics on each haplotype block in all populations (GWD, MSL, YRI, ESN, LWK) for each gene for all populations)
- Summary.HapStats.1kg.h3a.ESN.txt (Summary statistics on each haplotype block in each gene for ESN population)
- Summary.HapStats.1kg.h3a.GWD.txt (Summary statistics on each haplotype block in each gene for GWD population)
- Summary.HapStats.1kg.h3a.LWK.txt (Summary statistics on each haplotype block in each gene for LWK population)
- Summary.HapStats.1kg.h3a.MSL.txt (Summary statistics on each haplotype block in each gene for MSL population)
- Summary.HapStats.1kg.h3a.YRI.txt (Summary statistics on each haplotype block in each gene for YRI population)

Data are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

Acknowledgements

We thank members of the TrypanoGEN+ network for helpful comments. Members of the TrypanoGEN+ group of H3Africa are listed in www.trypanogen.net.

References

Abecasis GR, Auton A, Brooks LD, *et al.*: **An integrated map of genetic variation from 1,092 human genomes.** *Nature.* 2012; **491**(7422): 56–65.
PubMed Abstract | Publisher Full Text | Free Full Text

Abel L, Demenais F, Prata A, *et al.*: **Evidence for the segregation of a major gene in human susceptibility/resistance to infection by *Schistosoma mansoni*.** *Am J Hum Genet.* 1991; **48**(5): 959–970.
PubMed Abstract | Free Full Text

Adedoja A, Hoan NX, van Tong H, *et al.*: **Differential contribution of interleukin-10 promoter variants in malaria and schistosomiasis mono- and co-infections among Nigerian children.** *Trop Med Int Health.* 2018; **23**(1): 45–52.
PubMed Abstract | Publisher Full Text

Adedokun SA, Seamans BN, Cox NT, *et al.*: **Interleukin-4 and STAT6 promoter polymorphisms but not interleukin-10 or 13 are essential for schistosomiasis and associated disease burden among Nigerian children.** *Infect Genet Evol.* 2018; **65**: 28–34.
PubMed Abstract | Publisher Full Text

Adenowo AF, Oyinloye BE, Ogunyinka BI, *et al.*: **Impact of human schistosomiasis in sub-Saharan Africa.** *Braz J Infect Dis.* 2015; **19**(2): 196–205.
PubMed Abstract | Publisher Full Text

Antony JS, Ojurongbe O, Kremsner PG, *et al.*: **Lectin Complement Protein Collectin 11 (CL-K1) and Susceptibility to Urinary Schistosomiasis.** *PLoS Negl Trop Dis.* 2015; **9**(3): e0003647.
PubMed Abstract | Publisher Full Text | Free Full Text

Beck T, Shorter T, Brookes AJ: **GWAS Central: A comprehensive resource for the discovery and comparison of genotype and phenotype data from genome-wide association studies.** *Nucleic Acids Res.* 2020; **48**(D1): D933–D940.
PubMed Abstract | Publisher Full Text | Free Full Text

Beltrame A, Guerriero M, Angheben A, *et al.*: **Accuracy of parasitological and immunological tests for the screening of human schistosomiasis in immigrants and refugees from African countries: An approach with Latent Class Analysis.** *PLoS Negl Trop Dis.* 2017; **11**(6): e0005593.
PubMed Abstract | Publisher Full Text | Free Full Text

Casacuberta-Partal M, Hoekstra PT, Kornelis D, *et al.*: **An innovative and user-friendly scoring system for standardised quantitative interpretation of the urine-based point-of-care strip test (POC-CCA) for the diagnosis of intestinal schistosomiasis: A proof-of-concept study.** *Acta Trop.* 2019; **199**: 105150.
PubMed Abstract | Publisher Full Text

Colley DG, Andros TS, Campbell CH Jr: **Schistosomiasis is more prevalent than previously thought: What does it mean for public health goals, policies, strategies, guidelines and intervention programs?** *Infect Dis Poverty.* 2017; **6**(1): 63.
PubMed Abstract | Publisher Full Text | Free Full Text

Colley DG, Bustinduy AL, Secor WE, *et al.*: **Human schistosomiasis.** *Lancet.* 2014; **383**(9936): 2253–2264.
PubMed Abstract | Publisher Full Text | Free Full Text

Colley DG, King CH, Kittur N, *et al.*: **Evaluation, Validation, and Recognition of the Point-of-Care Circulating Cathodic Antigen, Urine-Based Assay for Mapping *Schistosoma mansoni* Infections.** *Am J Trop Med Hyg.* 2020; **103**(1_Suppl): 42–49.
PubMed Abstract | Publisher Full Text | Free Full Text

Coulibaly JT, N'Gbesso YK, Knopp S, *et al.*: **Accuracy of urine circulating cathodic antigen test for the diagnosis of *Schistosoma mansoni* in preschool-aged children before and after treatment.** *PLoS Negl Trop Dis.* 2013; **7**(3): e2109.
PubMed Abstract | Publisher Full Text | Free Full Text

Dawson EM, Sousa-Figueiredo JC, Kabatereine NB, *et al.*: **Intestinal schistosomiasis in pre school-aged children of Lake Albert, Uganda: Diagnostic accuracy of a rapid test for detection of anti-schistosome antibodies.** *Trans R Soc Trop Med Hyg.* 2013; **107**(10): 639–647.
PubMed Abstract | Publisher Full Text

Deelder AM, Klappe HT, van den Aardweg GJ, *et al.*: ***Schistosoma mansoni*: Demonstration of two circulating antigens in infected hamsters.** *Exp Parasitol.* 1976; **40**(2): 189–197.
PubMed Abstract | Publisher Full Text

Dessein H, Duflot N, Romano A, *et al.*: **Genetic algorithms identify individuals with high risk of severe liver disease caused by schistosomes.** *Hum Genet.* 2020; **139**(6–7): 821–831.
PubMed Abstract | Publisher Full Text

Ellis MK, Zhao ZZ, Chen HG, *et al.*: **Analysis of the 5q31 33 locus shows an association between single nucleotide polymorphism variants in the *IL-5* gene and symptomatic infection with the human blood fluke, *Schistosoma japonicum*.** *J Immunol.* 2007; **179**(12): 8366–8371.
PubMed Abstract | Publisher Full Text | Free Full Text

Eriksson J, Reimert CM, Kabatereine NB, *et al.*: **The 434(G>C) polymorphism within the coding sequence of Eosinophil Cationic Protein (ECP) correlates with the natural course of *Schistosoma mansoni* infection.** *Int J Parasitol.* 2007; **37**(12): 1359–1366.
PubMed Abstract | Publisher Full Text

Gatlin MR, Black CL, Mwinzi PN, *et al.*: **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.* 2009; **3**(2): e375.
PubMed Abstract | Publisher Full Text | Free Full Text

Glinz D, Silué KD, Knopp S, *et al.*: **Comparing diagnostic accuracy of Kato-Katz, Koga agar plate, ether-concentration, and FLOTAC for *Schistosoma mansoni* and soil-transmitted helminths.** *PLoS Negl Trop Dis.* 2010; **4**(7): e754.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Grant AV, Araujo MI, Ponte EV, *et al.*: **Polymorphisms in *IL10* are associated with total Immunoglobulin E levels and *Schistosoma mansoni* infection intensity in a Brazilian population.** *Genes Immun.* 2011; **12**(1): 46–50.
PubMed Abstract | **Publisher Full Text**

Grant AV, Araujo MI, Ponte EV, *et al.*: **Functional polymorphisms in *IL13* are protective against high *Schistosoma mansoni* infection intensity in a Brazilian population.** *PLoS One.* 2012; **7**(5): e35863.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Hamburger J, He-Na, Abbasi I, *et al.*: **Polymerase chain reaction assay based on a highly repeated sequence of Schistosoma haematobium: A potential tool for monitoring schistosome-infested water.** *Am J Trop Med Hyg.* 2001; **65**(6): 907–911.
PubMed Abstract | **Publisher Full Text**

He H, Isnard A, Kouriba B, *et al.*: **A *STAT6* gene polymorphism is associated with high infection levels in urinary schistosomiasis.** *Genes Immun.* 2008; **9**(3): 195–206.
PubMed Abstract | **Publisher Full Text**

Hirschhorn JN, Lohmueller K, Byrne E, *et al.*: **A comprehensive review of genetic association studies.** *Genet Med.* 2002; **4**(2): 45–61.
PubMed Abstract | **Publisher Full Text**

Idris ZM, Yazdanbakhsh M, Adegnika AA, *et al.*: **A Pilot Study on Cytotoxic T Lymphocyte-4 Gene Polymorphisms in Urinary Schistosomiasis.** *Genet Test Mol Biomarkers.* 2012; **16**(6): 488–492.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Isnard A, Kouriba B, Doumbo O, *et al.*: **Association of *rs7719175*, located in the *IL13* gene promoter, with *Schistosoma haematobium* infection levels and identification of a susceptibility haplotype.** *Genes Immun.* 2011; **12**(1): 31–39.
PubMed Abstract | **Publisher Full Text**

Kamdem SD, Moyou-Somo R, Brombacher F, *et al.*: **Host Regulators of Liver Fibrosis During Human Schistosomiasis.** *Front Immunol.* 2018; **9**: 2781.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Katz N, Chaves A, Pellegrino J: **A simple device for quantitative stool thick-smear technique in Schistosomiasis mansoni.** *Rev Inst Med Trop Sao Paulo.* 1972; **14**(6): 397–400.
PubMed Abstract

Kittur N, Castleman JD, Campbell CH, *et al.*: **Comparison of *Schistosoma mansoni* Prevalence and Intensity of Infection, as Determined by the Circulating Cathodic Antigen Urine Assay or by the Kato-Katz Fecal Assay: A Systematic Review.** *Am J Trop Med Hyg.* 2016; **94**(3): 605–610.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Kongs A, Marks G, Verlé P, *et al.*: **The unreliability of the Kato-Katz technique limits its usefulness for evaluating *S. mansoni* infections.** *Trop Med Int Health.* 2001; **6**(3): 163–169.
PubMed Abstract | **Publisher Full Text**

Kouriba B, Chevillard C, Bream JH, *et al.*: **Analysis of the 5q31-q33 Locus Shows an Association between IL13-1055C/T IL-13-591A/G Polymorphisms and *Schistosoma haematobium* Infections.** *J Immunol.* 2005; **174**(10): 6274–6281.
PubMed Abstract | **Publisher Full Text**

Laird NM, Horvath S, Xu X: **Implementing a unified approach to family-based tests of association.** *Genet Epidemiol.* 2000; **19 Suppl 1**: S36–S42.
PubMed Abstract | **Publisher Full Text**

Laird NM, Lange C: **The Role of Family-Based Designs in Genome-Wide Association Studies.** *Statist Sci.* 2009; **24**(4): 388–397.
Publisher Full Text

Larkin BM, Smith PM, Ponichtera HE, *et al.*: **Induction and regulation of pathogenic Th17 cell responses in schistosomiasis.** *Semin Immunopathol.* 2012; **34**(6): 873–888.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Li Y, Levran O, Kim J, *et al.*: **Extreme sampling design in genetic association mapping of quantitative trait loci using balanced and unbalanced case-control samples.** *Sci Rep.* 2019; **9**(1): 15504.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Loucoubar C, Grant AV, Bureau JF, *et al.*: **Detecting multi-way epistasis in family-based association studies.** *Brief Bioinform.* 2017; **18**(3): 394–402.
PubMed Abstract | **Publisher Full Text**

LoVerde PT: **Schistosomiasis.** *Adv Exp Med Biol.* 2019; **1154**: 45–70.
PubMed Abstract | **Publisher Full Text**

Marquet S, Abel L, Hillaire D, *et al.*: **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.* 1999; **7**(1): 88–97.
PubMed Abstract | **Publisher Full Text**

Marquet S, Abel L, Hillaire D, *et al.*: **Genetic localization of a locus controlling the intensity of infection by *Schistosoma mansoni* on chromosome 5q31-q33.** *Nat Genet.* 1996; **14**(2): 181–184.
PubMed Abstract | **Publisher Full Text**

Mayaka M: **Etude épidémiologique de la bilharziose à Schistosoma mansoni en milieu scolaire: Cas du groupement de Kiyanika.** [Thèse de Doctorat]. Université de Kongo. 2001.
Reference Source

Mbow M, Larkin BM, Meurs L, *et al.*: **T-Helper 17 Cells Are Associated With Pathology in Human Schistosomiasis.** *J Infect Dis.* 2013; **207**(1): 186–195.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

McManus DP, Dunne DW, Sacko M, *et al.*: **Schistosomiasis.** *Nat Rev Dis Primers.* 2018; **4**(1): 13.
PubMed Abstract | **Publisher Full Text**

Mewamba EM, Nyangiri OA, Noyes HA, *et al.*: **The Genetics of Human Schistosomiasis Infection Intensity and Liver Disease: A Review.** *Front Immunol.* 2021a; **12**: 613468.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Mewamba EM, Tiofack AAZ, Kamdem CN, *et al.*: **Field assessment in Cameroon of a reader of POC-CCA lateral flow strips for the quantification of *Schistosoma mansoni* Circulating Cathodic Antigen in urine.** *PLoS Negl Trop Dis.* 2021b; **15**(7): e0009569.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Mnkugwe RH, Minzi OS, Kinung'hi SM, *et al.*: **Prevalence and correlates of intestinal schistosomiasis infection among school-aged children in North-Western Tanzania.** *PLoS One.* 2020; **15**(2): e0228770.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Montresor A, World Health Organization: **Helminth control in school-age children: A guide for managers of control programmes.** (2nd ed). World Health Organization. 2011.
Reference Source

Müller-Myhsok B, Stelma FF, Guissé-Sow F, *et al.*: **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.* 1997; **61**(2): 452–454.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Nash TE, Deelder AM: **Comparison of Four Schistosome Excretory-Secretory Antigens: Phenol Sulfuric Test Active Peak, Cathodic Circulating Antigen, Gut-Associated Proteoglycan, and Circulating Anodic Antigen.** *Am J Trop Med Hyg.* 1985; **34**(2): 236–241.
PubMed Abstract | **Publisher Full Text**

Noyes H: **Consent forms and CRFs for family based or case control schisto studies.** Harvard Dataverse, V1. 2021a.
http://www.doi.org/10.7910/DVN/6OLLGW

Noyes H: **Haplotype Genotyping.** [Data set]. Harvard Dataverse. 2021b.
http://www.doi.org/10.7910/DVN/OVQPCO

Ott J, Kamatani Y, Lathrop M: **Family-based designs for genome-wide association studies.** *Nat Rev Genet.* 2011; **12**(7): 465–474.
PubMed Abstract | **Publisher Full Text**

Ouf EA, Ojurongbe O, Akindele AA, *et al.*: **Ficolin-2 Levels and *FCN2* Genetic Polymorphisms as a Susceptibility Factor in Schistosomiasis.** *J Infect Dis.* 2012; **206**(4): 562–570.
PubMed Abstract | **Publisher Full Text**

Panarella M, Burkett KM: **A Cautionary Note on the Effects of Population Stratification Under an Extreme Phenotype Sampling Design.** *Front Genet.* 2019; **10**: 398.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Patel DD, Kuchroo VK: **Th17 Cell Pathway in Human Immunity: Lessons from Genetics and Therapeutic Interventions.** *Immunity.* 2015; **43**(6): 1040–1051.
PubMed Abstract | **Publisher Full Text**

Saotoing P, Vroumsia T, Njan AM, *et al.*: **Epidemiological survey of schistosomiasis due to *Schistosoma haematobium* in some primary schools in the town of Maroua, Far North Region Cameroon.** *Int J Trop Med.* 2011; **6**(2): 19–24.
Publisher Full Text

Sousa MS, van Dam GJ, Pinheiro MCC, *et al.*: **Performance of an Ultra-Sensitive Assay Targeting the Circulating Anodic Antigen (CAA) for Detection of *Schistosoma mansoni* Infection in a Low Endemic Area in Brazil.** *Front Immunol.* 2019; **10**: 682.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Spielman RS, McGinnis RE, Ewens WJ: **Transmission Test for Linkage Disequilibrium: The Insulin Gene Region and Insulin-dependent Diabetes Mellitus (IDDM).** *Am J Hum Genet.* 1993; **52**(3): 506–516.
PubMed Abstract | **Free Full Text**

The H3Africa Consortium, Rotimi C, Abayomi A, *et al.*: **Research capacity. Enabling the genomic revolution in Africa.** *Science.* 2014; **344**(6190): 1346–1348.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Tiongco RE, Paragas NA, Dominguez MJ, *et al.*: **ABO blood group antigens may be associated with increased susceptibility to schistosomiasis: A systematic review and meta-analysis.** *J Helminthol.* 2018; **94**: e21.
PubMed Abstract | **Publisher Full Text**

van Dam GJ, de Dood CJ, Lewis M, *et al.*: **A robust dry reagent lateral flow assay for diagnosis of active schistosomiasis by detection of *Schistosoma* circulating anodic antigen.** *Exp Parasitol.* 2013; **135**(2): 274–282.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

van Dam GJ, Wichers JH, Ferreira TMF, *et al.*: **Diagnosis of schistosomiasis by reagent strip test for detection of circulating cathodic antigen.** *J Clin Microbiol.* 2004; **42**(12): 5458–5461.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

WHO: **Enhancing implementation of schistosomiasis control and elimination programmes.** 2012.
Reference Source

WHO: **Schistosomiasis, progress report 2001–2011, strategic plan 2012–2020.** World Health Organization. 2013.
Reference Source

World Health Organisation: **Kato-Katz technique—Cellophane faecal thick smear.** 1994.
Reference Source

World Health Organisation (Ed.): **Preventive chemotherapy in human helminthiasis: Coordinated use of anthelminthic drugs in control interventions ; a manual for health professionals and programme managers**. World Health Organization. 2006.
Reference Source

Zinn-Justin A, Marquet S, Hillaire D, *et al.*: **Genome search for additional human loci controlling infection levels by Schistosoma mansoni.** *Am J Trop Med Hyg.* 2001; **65**(6): 754–758.
PubMed Abstract | **Publisher Full Text**

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 2

Reviewer Report 21 February 2022

https://doi.org/10.21956/aasopenres.14458.r29061

© **2022 Armoo S.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✔ **Samuel K. Armoo**
Biomedical and Public Health Research Unit, Water Research Institute, Council for Scientific and Industrial Research, Accra, Ghana

Approved.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 20 October 2021

https://doi.org/10.21956/aasopenres.14319.r28888

© **2021 Armoo S.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

? **Samuel K. Armoo**
Biomedical and Public Health Research Unit, Water Research Institute, Council for Scientific and Industrial Research, Accra, Ghana

Summary:
This manuscript describes the relationship between underlying human host genotypes and the level of susceptibility or otherwise to high *Schistosoma mansoni* worm burdens. The authors tested 24 candidate genes in four endemic African countries and found an over-representation of some

haplotypes in study participants with high worm burdens. This manuscript falls within the scope of this journal, but there should be some adjustments to make it suitable for indexing.

Specific comments:

- Table 1: Provide any amino acid changes from SNPs and their proposed functional effects.

Sampling procedure:

- **Inclusion Criteria**: Specify whether the duration of stay in the community is an inclusion criterion. If not, why? In addition, does the child need to be born in the community? If not, why?

Diagnosis:

Kato-Katz:

1. Which quality control procedures were used for the Kato-Katz test?

2. Why not use *S. mansoni* eggs per gram of stool for quantification of worm burdens?

CCA:

1. Armoo *et al.* (2020[1]) detected low specificity of the POC-CCA assay, when trace bands were included in high transmission areas in Ghana. Specificity increased significantly (from 12.9 to 76.5%) when trace bands were excluded. So kindly consider the role of trace bands in this study.

2. CCA should be complemented by Kato-Katz or qPCR to determine worm burden phenotype.

3. Which quality control procedures were used for the CCA and CAA assays?

References

1. Armoo S, Cunningham LJ, Campbell SJ, Aboagye FT, et al.: Detecting Schistosoma mansoni infections among pre-school-aged children in southern Ghana: a diagnostic comparison of urine-CCA, real-time PCR and Kato-Katz assays.*BMC Infect Dis*. 2020; **20** (1): 301 PubMed Abstract | Publisher Full Text

Is the rationale for, and objectives of, the study clearly described?

Yes

Is the study design appropriate for the research question?

Yes

Are sufficient details of the methods provided to allow replication by others?

No

Are the datasets clearly presented in a useable and accessible format?

Not applicable

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Medical parasitology and molecular genetics of neglected tropical diseases.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 02 Dec 2021

Oscar Nyangiri, Makerere University, Kampala, Uganda

Summary:

This manuscript describes the relationship between underlying human host genotypes and the level of susceptibility or otherwise to high *Schistosoma mansoni* worm burdens. The authors tested 24 candidate genes in four endemic African countries and found an over-representation of some haplotypes in study participants with high worm burdens. This manuscript falls within the scope of this journal, but there should be some adjustments to make it suitable for indexing.

We clarify that this is a protocol describing how we will test these 24 candidate genes and that we do not yet have results.

Specific comments:

- **Table 1: Provide any amino acid changes from SNPs and their proposed functional effects.**

We used the tool variant effect predictor here http://grch37.ensembl.org/info/docs/tools/vep/ to query the SNP IDs (rsIDs). In the legend of table 1 we add a reference to supplementary Supplementary_tableVEP1.xls (Noyes, 2021b)

Amino acid changes are under column header “Amino_acids” while functional effects are under column headers

“Consequence”, which gives the position whether upstream, downstream of gene, regulatory, in gene whether missense or synonymous

“BIOTYPE” which allows for a check whether the variant is in protein coding region or not

“SIFT” and “PolyPhen” which gives a score of whether the variant is tolerated, benign or deleterious,

“CLIN_SIG” which is clinical significance corresponding to whether the variant is a risk factor or protective

Sampling procedure:

- **Inclusion Criteria: Specify whether the duration of stay in the community is an inclusion criterion. If not, why? In addition, does the child need to be born in the community? If not, why?**

Thanks for pointing out this omission. We aim to recruit members resident in the communities and therefore duration of stay should be in the last 1 year. We have added this under the section “inclusion criteria”

Additionally, members born in the community are bona fide community members and should therefore be included.

In summary, the rationale for adding these criteria: To avoid spurious results from imported infections or uninfected individuals, the members should have homogenous exposure to the risk in this community. Therefore, they should either be born in the community or if not; have spent a considerable amount of time in the community to be considered members with the same exposure as the rest of the community members.

Diagnosis:

Kato-Katz:

1. **Which quality control procedures were used for the Kato-Katz test?**

We will homogenize stool by mixing in a polythene bag to make the distribution more uniform. After this we will pick the stool using spatula from different areas of the stool sample and make the Kato Katz stool slide. We will take duplicate slides and have two separate readings of egg counts by two technicians. The average of these readings is then taken as the eggs per gram.

2. Why not use *S. mansoni* eggs per gram of stool for quantification of worm burdens?

We will indeed use two parallel tests: KK eggs per gram and POC CCA. CAA will also be used for all Uganda samples. Although they do not measure the exact worm/egg antigens or actual eggs observed, there is some concordance between these tests as published for CCA and KK (Coulibaly et al., 2013; Dawson et al., 2013; Kittur et al., 2016) and for KK and CAA (van Dam et al., 2013). The order of sensitivity is CAA>CCA>KK. Since our study is primarily interested in the top 20% of worm burden, the most sensitive tests, reader based POC-CCA and CAA are chosen to maximize the ability to discriminate between samples and allow a better estimation of genotypes over-represented in extreme phenotypes of the top 20% worm burden

In the introduction under the section of how the proposed study aims to improve on previous studies, we state that CAA/CCA are more reproducible than the Kato-Katz method. "The Kato-Katz method of counting eggs in stool samples is known to lack sensitivity and this is compounded by day to day variation in egg excretions and uneven distribution of eggs across an individual stool (Kongs et al., 2001)."

CCA:

1. **Armoo *et al.* (2020[1]) detected low specificity of the POC-CCA assay, when trace bands were included in high transmission areas in Ghana. Specificity increased significantly (from 12.9 to 76.5%) when trace bands were excluded. So kindly consider the role of trace bands in this study.**

Thanks for this observation. We note from Armoo et al. (2020[1]) that inclusion of trace bands reduced specificity. The gold standard to assess specificity in Armoo et al, 2020 was the combination of POC-CCA and real time PCR Taqman.

To account for this apparent lack of specificity from POC-CCA trace bands; our study uses a POC-CCA reader to convert the signal from the all bands in the POC-CCA test to optical intensity (see section "Using the RDT-Reader for quantitative POC-CCA"). In this way, we have a continuity from barely positive to highly positive. The lower limit of detection on the POC-CCA reader is 40mV of reflected optical density, which detects very faint bands. We consider these trace bands in the bottom 20% of worm burden.

We also use CAA the most sensitive method as a gold standard to establish whether CCA

trace was actually positive. Our CAA assay has a sensitivity of up to 30 pg/mL of the circulating anodic antigen.
We are also sampling in high transmission areas. According to Colley et al. 2020 trace bands are generally true positives in high transmission areas but are more likely to be false positives in low transmission areas
It is important to note in the family-based studies that we are only genotyping the individuals who have the heaviest 20% of worm burdens and none of these will have trace CCA readings.

2. CCA should be complemented by Kato-Katz or qPCR to determine worm burden phenotype.

In the section "Diagnosis" we describe all the methods we will use for diagnosis of schistosomiasis. Here, we describe Kato-Katz, CCA, a reader based CCA assay and CAA. We have therefore complemented CCA with KK.
The CAA test will be done for a subset of those tested by both KK and reader based CCA assay to check for concordance between these tests.
At least one of either KK or CAA have been used previously for similar analyses of genetic conditioning of egg or worm burden phenotype (Marquet et al., 1996; Marquet et al., 1999; Müller-Myhsok et al., 1997; Zinn-Justin et al., 2001). Mewamba et al 2021 has also shown how the reader-based method correlates with KK and can be used quantitatively.
Regarding qPCR, it is a helpful additional method but at this time we are not able to do qPCR because of resource constraints.

3. Which quality control procedures were used for the CCA and CAA assays?

The CCA was done in two ways

1. 1. scoring the reading by eye from negative (0), trace (1), intermediate (2) and strong (3) as described in "Circulating Cathodic antigen (CCA) test" and using the G-scores as a standardization scoring system (as explained in Mewamba et al 2021)
   2. Using the reader to pick the optical intensity of the bands as described in section "Using the RDT-Reader for quantitative POC-CCA"

In our analysis we will compare the two readings, where we expect to see a concordance between the qualitative readings by eye with those from the reader. We have observed this already in one site, Cameroon (Mewamba et al 2021).
We will use CAA on a subset of samples to validate our CCA results. CAA comes with standards from the manufacturer which we use to calibrate our readings.
In addition to these, we will compare our readings with KK to check for trends. Particularly we will be interested to see if top/bottom samples in CAA are also matching respective top/bottom samples in CCA and KK.

References

1. Armoo S, Cunningham LJ, Campbell SJ, Aboagye FT, et al.: Detecting Schistosoma mansoni infections among pre-school-aged children in southern Ghana: a diagnostic comparison of urine-CCA, real-time PCR and Kato-Katz assays.*BMC Infect Dis*. 2020; **20** (1): 301 PubMed Abstract | Publisher Full Text
2. Colley DG, King CH, Kittur N, Ramzy RMR, Secor WE, Fredericks-James M, Ortu G, Clements MN, Ruberanziza E, Umulisa I, Wittmann U, Campbell CH. Evaluation, Validation, and Recognition of the Point-of-Care Circulating Cathodic Antigen, Urine-

Based Assay for Mapping *Schistosoma mansoni* Infections. Am J Trop Med Hyg. 2020 Jul;103(1_Suppl):42-49. doi: 10.4269/ajtmh.19-0788. PMID: 32400347; PMCID: PMC7351311.

3. Coulibaly, Jean T., Yves K. N'Gbesso, Stefanie Knopp, Nicaise A. N'Guessan, Kigbafori D. Silué, Govert J. van Dam, Eliézer K. N'Goran, and Jürg Utzinger. 'Accuracy of Urine Circulating Cathodic Antigen Test for the Diagnosis of Schistosoma Mansoni in Preschool-Aged Children before and after Treatment'. *PLoS Neglected Tropical Diseases* 7, no. 3 (2013): e2109. https://doi.org/10.1371/journal.pntd.0002109.
4. Dawson, Emily M., José C. Sousa-Figueiredo, Narcis B. Kabatereine, Michael J. Doenhoff, and J. Russell Stothard. 'Intestinal Schistosomiasis in Pre School-Aged Children of Lake Albert, Uganda: Diagnostic Accuracy of a Rapid Test for Detection of Anti-Schistosome Antibodies'. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 107, no. 10 (October 2013): 639–47. https://doi.org/10.1093/trstmh/trt077.
5. Kittur, Nupur, Jennifer D. Castleman, Carl H. Campbell, Charles H. King, and Daniel G. Colley. 'Comparison of Schistosoma Mansoni Prevalence and Intensity of Infection, as Determined by the Circulating Cathodic Antigen Urine Assay or by the Kato-Katz Fecal Assay: A Systematic Review'. *The American Journal of Tropical Medicine and Hygiene* 94, no. 3 (March 2016): 605–10. https://doi.org/10.4269/ajtmh.15-0725.
6. Kongs, A., G. Marks, P. Verlé, and P. Van der Stuyft. 'The Unreliability of the Kato-Katz Technique Limits Its Usefulness for Evaluating S. Mansoni Infections'. *Tropical Medicine & International Health* 6, no. 3 (2001): 163–69. https://doi.org/10.1046/j.1365-3156.2001.00687.x.
7. Marquet, S., L. Abel, D. Hillaire, H. Dessein, J. Kalil, J. Feingold, J. Weissenbach, and A. J. Dessein. 'Genetic Localization of a Locus Controlling the Intensity of Infection by *Schistosoma Mansoni* on Chromosome 5q31-Q33'. *Nature Genetics* 14, no. 2 (October 1996): 181–84. https://doi.org/10.1038/ng1096-181.
8. Marquet, S., L. Abel, D. Hillaire, and A. Dessein. 'Full Results of the Genome-Wide Scan Which Localises a Locus Controlling the Intensity of Infection by *Schistosoma Mansoni* on Chromosome 5q31-Q33'. *European Journal of Human Genetics: EJHG* 7, no. 1 (January 1999): 88–97. https://doi.org/10.1038/sj.ejhg.5200243.
9. Müller-Myhsok, B., F. F. Stelma, F. Guissé-Sow, B. Muntau, T. Thye, G. D. Burchard, B. Gryseels, and R. D. Horstmann. 'Further Evidence Suggesting the Presence of a Locus, on Human Chromosome 5q31-Q33, Influencing the Intensity of Infection with Schistosoma Mansoni'. *American Journal of Human Genetics* 61, no. 2 (August 1997): 452–54. https://doi.org/10.1016/S0002-9297(07)64073-7.
10. Zinn-Justin, A., S. Marquet, D. Hillaire, A. Dessein, and L. Abel. 'Genome Search for Additional Human Loci Controlling Infection Levels by *Schistosoma Mansoni*'. *The American Journal of Tropical Medicine and Hygiene* 65, no. 6 (December 2001): 754–58. https://doi.org/10.4269/ajtmh.2001.65.754.
11. Mewamba, Estelle Mezajou, Arnol Auvaker Zebaze Tiofack, Cyrille Nguemnang Kamdem, Romuald Isaka Kamwa Ngassam, Mureille Carole Tchami Mbagnia, Oscar Nyangiri, Harry Noyes, Hilaire Marcaire Womeni, Flobert Njiokou, and Gustave Simo. 'Field Assessment in Cameroon of a Reader of POC-CCA Lateral Flow Strips for the Quantification of Schistosoma Mansoni Circulating Cathodic Antigen in Urine'. *PLoS Neglected Tropical Diseases* 15, no. 7 (July 2021): e0009569. https://doi.org/10.1371/journal.pntd.0009569.

Competing Interests: The authors declare no competing interests

Reviewer Report 23 August 2021

https://doi.org/10.21956/aasopenres.14319.r28821

© 2021 Kamdem S. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Severin Donald Kamdem
Departement of pathology, School of Medicine, University of Utah, Salt Lake City, UT, USA

The manuscript *Candidate gene family-based and case-control studies of susceptibility to high Schistosoma mansoni worm burden in African children: a protocol [version 1; peer review: awaiting peer review]* by Nyangiri *et al.* aims to test 24 candidate genes, mainly in the Th2 and Th17 pathways, for association with *S. mansoni* infection intensity in four African countries (Ivory Coast, Cameroon, Uganda and Democratic Republic of Congo), using family based and case-control approaches. Regarding the dissimilarity observed in Schistosomiasis infection and related morbidity, this study is relevant for the field and could help for a better management of Schistosomiasis in endemic areas.

The authors have done a preliminary work by reviewing all genetic factors that control Schistosomiasis infection (Mewamba *et al.* 2021, Frontiers in Immunology) and seem to have the baseline knowledge to conduct the proposed study. The objectives of the study are clearly established and the description of the Methodology appropriate. The respect of the ethic is well described and the authorization needed gathered. The proposed statistics and meta-analysis are comprehensive.

However, the following suggestions will help improve the overall quality of the results.

Since the authors intend to include school children, they will consider obtaining authorization from Ministries of Education.

In addition to CCA and CAA, the authors should consider taking several stool samples (2 to 3) from consecutive days for KK analysis for an increased sensitivity using the resulting geometric mean of burden intensity. In fact, the CCA test could easily lead to false positive results in endemic areas, and the CAA is a good correlate for worm burden and not of eggs burden (as CAA levels indeed reflect dynamics of worm burdens and not of egg excretion or worm fecundity). KK remains the appropriate technique to accurately determine the egg burden. Of note, Schistosomiasis pathology is eggs (and not worms) driven. CCA and CAA can be useful in determining controls (true negative with all techniques) in the case-controls studies in Uganda.

The authors propose to use the top 25% positive participants as participants with high burden.

Considering the WHO classification for *S. mansoni* infection (light < 100; 100 < Moderate <400 or heavy > 400), what would be the contingency plan if participants in the top 25% have a burden lower than 400 EPG? Or 100EPG?

The Figure 1 is quite confusing. The authors should consider revising it. Also, excluded samples should be mentioned. In addition, participants should be screened for clinical signs of anemia and excluded prior to blood collection.

The authors should consider screening for other confections such as malaria etc. - generally co-endemic in these regions, as co-infections could affect the susceptibility of Schistosomiasis infection. Risk factors of susceptibility/Resistance to Schistosomiasis infection will be valuable information to have.

It is not mentioned if the participants will be treated with Praziquantel at the end of the study.

For the case-control study, the authors should correct for socio-demographic data to ensure minimal variation across the groups. For example, to fully attribute the difference in infection intensity to the genetics, the authors should make sure factors like age; gender; BMI etc. are not different between cases and controls. This will aid in minimizing confounding factors.

Is the rationale for, and objectives of, the study clearly described?
Yes

Is the study design appropriate for the research question?
Yes

Are sufficient details of the methods provided to allow replication by others?
Partly

Are the datasets clearly presented in a useable and accessible format?
Not applicable

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Infectious Diseases and Fibrosis

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 15 Nov 2021

Oscar Nyangiri, Makerere University, Kampala, Uganda

We appreciate this review and the useful comments received which will help to improve our manuscript. Below, we seek to address each of the concerns; with concerns and in bold and response in normal type font.

Since the authors intend to include school children, they will consider obtaining

authorization from Ministries of Education.
We considered the ethical procedures in each country, and in the section "Ethical approval and consent to participate" we describe what each country is doing which has been accepted by the ethical review committee. We already include district level school authorities in our authorization and community engagement work. We agree that the suggestion to involve the Ministry of Education is useful and we will suggest to our community engagement team to include the Ministry of Education where it is not currently included. Therefore in this section we add the statement "In all countries, since schools are involved, we will consult the ministry of education in conducting the study."

In addition to CCA and CAA, the authors should consider taking several stool samples (2 to 3) from consecutive days for KK analysis for an increased sensitivity using the resulting geometric mean of burden intensity. In fact, the CCA test could easily lead to false positive results in endemic areas, and the CAA is a good correlate for worm burden and not of eggs burden (as CAA levels indeed reflect dynamics of worm burdens and not of egg excretion or worm fecundity). KK remains the appropriate technique to accurately determine the egg burden. Of note, Schistosomiasis pathology is eggs (and not worms) driven. CCA and CAA can be useful in determining controls (true negative with all techniques) in the case-controls studies in Uganda.
Although we agree that eggs rather than worms cause the fibrosis related pathology the correlation between egg burden and fibrosis is weak (Kamden et al. 2018). We will be looking for genetic associations with worm burden and we have selected genes that have been or could be associated with either egg or worm burden and therefore a test that directly measures this burden is more appropriate than a proxy such as egg count. A candidate gene study of genetic associations with pathology would include a very different set of genes
In the current study we also conduct at least one KK to check the correlation with both CAA and CCA tests. This is what is practical in the field for us. The CAA test is a true quantitative measure (van Dam et al., 2013), while the reflection from the CCA test has been modified for use as a quantitative test (Mewamba et al., 2021).
Whilst we accept that multiple KK readings on consecutive days gives a more reproducible results than duplicate readings on a single day, we are planning to use the CCA as it is more reproducible from a single reading than the KK.
Kato Katz, the current standard method for detecting the egg burden requires 3 different readings from a non-homogenous stool sample. It also depends to a great extent the time of sampling. For these practical reasons, we will base worm burden quantification on CAA and reader assisted CCA.

The authors propose to use the top 25% positive participants as participants with high burden. Considering the WHO classification for *S. mansoni* infection (light < 100; 100 < Moderate <400 or heavy > 400), what would be the contingency plan if participants in the top 25% have a burden lower than 400 EPG? Or 100EPG?
The egg burdens vary by community, where some communities will have predominantly light burdens or heavy burdens and these burdens will depend on the intensity of exposure and the history of MDA in the region. In so far as intensity of infection is controlled by genetic factors, we assume that a person with a worm burden in the top 20% in a community with low intensities of infection would be likely to be also in the top 20% in a

community with heavy worm burdens and therefore the absolute intensity of infection is not relevant. We therefore use relative burdens, where we define the highest burden as those over the 80th percentile of worm burdens. These are are refered to as the top 20% in the protocol. In the case control study, we compare the top 25% with the bottom 25%. With this definition, we therefore do not use the WHO definitions of light, moderate and heavy egg burdens.

The Figure 1 is quite confusing. The authors should consider revising it. Also, excluded samples should be mentioned. In addition, participants should be screened for clinical signs of anemia and excluded prior to blood collection.
We have revised this figure as shown in Figure1_flow_revised.docx

The authors should consider screening for other confections such as malaria etc. - generally co-endemic in these regions, as co-infections could affect the susceptibility of Schistosomiasis infection.
This is an important consideration. We note other helminth infections when screening samples by the KK test but these have been rare so far (<5%). We are aware of evidence that schistosome infections can modify malaria infections and drug response but there seems to be less work on the effect of malaria in schistosomiasis. We will screen for malaria infections if possible and include this as a covariate in the analysis. In the case control study, we actually screened for fever, diarrhea and infections with other parasitic worms as described in section “Additional data recorded per participant”.

Risk factors of susceptibility/Resistance to Schistosomiasis infection will be valuable information to have.
We have a questionnaire that seeks to capture environmental and demographic risk factors for schistosomiasis. In section “Data management” we stated that “Data will be collected in standardised forms as shown in supplementary data.”An example is the Uganda case control study case report form. Among these are included important variables like age, gender and sanitation facilities.

It is not mentioned if the participants will be treated with Praziquantel at the end of the study.
We confirm that participants will be treated with Praziquantel. These studies are nested in the National Control Programme, where we sample immediately before treatment with Praziquantel. We stated in the section “Screening procedures” that “All children will be treated with Praziquantel by the national Ministry of Health Mass Drug Administration teams according to the recommendation of the World Health Organization (World Health Organisation, 2006).”

For the case-control study, the authors should correct for socio-demographic data to ensure minimal variation across the groups. For example, to fully attribute the difference in infection intensity to the genetics, the authors should make sure factors like age; gender; BMI etc. are not different between cases and controls. This will aid in minimizing confounding factors.
Yes, this is actually the analysis plan. We mention we control for these factors in section Study 2: Case control association study under subtopic “Data analysis” we mention that

"Age, sex, MUAC and BMI will also be used as additional covariates" in our analysis.
MUAC is mid upper arm circumference
BMI Body mass index

REFERENCES:
van Dam GJ, de Dood CJ, Lewis M, *et al.*: **A robust dry reagent lateral flow assay for diagnosis of active schistosomiasis by detection of *Schistosoma* circulating anodic antigen.** *Exp Parasitol.* 2013; **135**(2): 274–282.
van Dam GJ, de Dood CJ, Lewis M, Deelder AM, van Lieshout L, Tanke HJ, van Rooyen LH, Corstjens PL. A robust dry reagent lateral flow assay for diagnosis of active schistosomiasis by detection of Schistosoma circulating anodic antigen. Exp Parasitol. 2013 Oct;135(2):274-82. doi: 10.1016/j.exppara.2013.06.017. Epub 2013 Jul 10. PMID: 23850995; PMCID: PMC3797160.

Kamdem, Severin Donald, Roger Moyou-Somo, Frank Brombacher, and Justin Komguep Nono. 2018. 'Host Regulators of Liver Fibrosis During Human Schistosomiasis'. *Frontiers in Immunology* 9: 2781. https://doi.org/10.3389/fimmu.2018.02781.

Mewamba EM, Tiofack AAZ, Kamdem CN, Ngassam RIK, Mbagnia MCT, Nyangiri O, Noyes H, Womeni HM, Njiokou F, Simo G. Field assessment in Cameroon of a reader of POC-CCA lateral flow strips for the quantification of Schistosoma mansoni circulating cathodic antigen in urine. PLoS Negl Trop Dis. 2021 Jul 14;15(7):e0009569. doi: 10.1371/journal.pntd.0009569. PMID: 34260610; PMCID: PMC8312929.

World Health Organisation, ed. *Preventive Chemotherapy in Human Helminthiasis: Coordinated Use of Anthelminthic Drugs in Control Interventions ; a Manual for Health Professionals and Programme Managers*. Geneva: World Health Organization, 2006.

Competing Interests: Authors declare no competing interests.