

# Spatial and Genetic Epidemiology of Hookworm in a Rural Community in Uganda

Rachel L. Pullan<sup>1\*</sup>, Narcis B. Kabatereine<sup>2</sup>, Rupert J. Quinnell<sup>3</sup>, Simon Brooker<sup>1,4</sup>

**1** Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom, **2** Vector Control Division, Ministry of Health, Kampala, Uganda, **3** Institute of Integrative and Comparative Biology, University of Leeds, Leeds, United Kingdom, **4** Malaria Public Health and Epidemiology Group, Kenya Medical Research Institute-Wellcome Trust Research Programme, Nairobi, Kenya

## Abstract

There are remarkably few contemporary, population-based studies of intestinal nematode infection for sub-Saharan Africa. This paper presents a comprehensive epidemiological analysis of hookworm infection intensity in a rural Ugandan community. Demographic, kinship, socioeconomic and environmental data were collected for 1,803 individuals aged six months to 85 years in 341 households in a cross-sectional community survey. Hookworm infection was assessed by faecal egg count. Spatial variation in the intensity of infection was assessed using a Bayesian negative binomial spatial regression model and the proportion of variation explained by host additive genetics (heritability) and common domestic environment was estimated using genetic variance component analysis. Overall, the prevalence of hookworm was 39.3%, with the majority of infections (87.7%) of light intensity ( $\leq 1000$  eggs per gram faeces). Intensity was higher among older individuals and was associated with treatment history with anthelmintics, walking barefoot outside the home, living in a household with a mud floor and education level of the household head. Infection intensity also exhibited significant household and spatial clustering: the range of spatial correlation was estimated to be 82 m and was reduced by a half over a distance of 19 m. Heritability of hookworm egg count was 11.2%, whilst the percentage of variance explained by unidentified domestic effects was 17.8%. In conclusion, we suggest that host genetic relatedness is not a major determinant of infection intensity in this community, with exposure-related factors playing a greater role.

**Citation:** Pullan RL, Kabatereine NB, Quinnell RJ, Brooker S (2010) Spatial and Genetic Epidemiology of Hookworm in a Rural Community in Uganda. *PLoS Negl Trop Dis* 4(6): e713. doi:10.1371/journal.pntd.0000713

**Editor:** Jeffrey Michael Bethony, George Washington University, United States of America

**Received:** December 8, 2009; **Accepted:** April 23, 2010; **Published:** June 15, 2010

**Copyright:** © 2010 Pullan 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 author and source are credited.

**Funding:** Rachel L. Pullan is supported by a Medical Research Council DTA-funded studentship (<http://www.mrc.ac.uk/consumption/groups/public/documents/content/mrc002630.pdf>). Simon Brooker is supported by a Research Career Development Fellowship from the Wellcome Trust (#081673) (<http://www.wellcome.ac.uk/Funding/Biomedical-science/Grants/Fellowships-and-personal-awards/Basic-biomedical-fellowships/WTD004431.htm>). The fieldwork was supported through Simon Brooker's fellowship. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: [rachel.pullan@lshtm.ac.uk](mailto:rachel.pullan@lshtm.ac.uk)

## Introduction

Recent years have seen an unprecedented expansion in financial and technical support for school-based deworming, with an increasing number of countries in sub-Saharan Africa implementing nationwide control. The initiation of such control does not however signal the end of epidemiological research and detailed data on patterns and risk factors for infection are still required for the refinement of ongoing control activities. It is surprising therefore that there are so few population-based studies of the epidemiology of hookworm in Africa. Those that do exist typically describe age-related changes in infection prevalence and intensity, demonstrating consistent increases with age, peaking in adults [1–6], pronounced aggregation of high intensity infection within high risk individuals [1,2], and villages [5,6], and providing evidence for predisposition to low or high intensity infection [1] using faecal egg counts as a proxy; only one study has looked at actual worm burdens [7]. Remarkably few investigations have been able to address spatial and/or genetic determinants of infection within African communities. However, our ability to investigate such factors today has been greatly facilitated by the increasing availability of high resolution spatial data and the advent of powerful analytical tools, including

Geographic Information Systems (GIS), readily implemented Bayesian analysis and quantitative genetic approaches [8–12]. For example, GIS-based studies in a cohort of South African primary school children have demonstrated considerable spatial clustering of infection within a smaller area, strongly influenced by several environmental factors [10]. Similarly, several studies outside Africa have reported a significant role for both host genetics and the family environment in determining hookworm infection intensity [12–14], although the only genetic epidemiology investigation of hookworm in Africa did not account for the effects of shared family environment [15].

This paper reports results from a population-based study of hookworm infection in Uganda. Our aim is to provide a detailed description of hookworm epidemiology for sub-Saharan Africa. In order to comprehensively investigate determinants of infection intensity two complementary analytical approaches are employed. First, negative binomial spatial modelling is used to investigate spatial variation in the intensity of infection, whilst adjusting for individual- and household-level covariates. Second, genetic variance component analysis, a quantitative genetic approach which takes into account familial relationships within and between households, is used to determine the relative contributions of host genetics (and other factors) to variation in infection intensity.

## Author Summary

Detailed descriptions of the epidemiology of intestinal nematode infections within affected communities remain of considerable importance for the effective design of disease control programmes. We therefore conducted a parasitological survey of 1,803 individuals living in a rural community in eastern Uganda. Two complementary analytical approaches were used to evaluate causes of variation in intensity of hookworm infection in this community, both of which exploit correlation structures between individuals. Infection intensity was higher among older individuals and was also associated with factors influencing environmental exposure to infection, socio-economic indicators and treatment history. After accounting for these risk factors, spatial correlation remained evident between households less than 82m apart. Focusing further on similarities within households, our results suggest that 11% of variation in infection intensity could be explained by genetic differences between individuals and 18% by unmeasured factors associated with the domestic environment. Taken together, these results suggest that exposure-related factors, rather than host genetics, have the greatest influence on variation in infection intensities.

## Methods

### Study area and population

The study was conducted in 2008 in four villages in Mulanda sub-county, which is located in Tororo district, eastern Uganda. The area is characterised by dry savannah grassland interrupted by bare rocky outcrops and lower lying swamps, with an average daytime temperature of 27°C and two rainy seasons (March to May and August to October). The majority of inhabitants are involved in rural subsistence farming, with garden plots in very close proximity to farmers' compounds. Houses are predominantly of traditional construction (mud walls, thatched roofs), approximately 95% of households have a pit latrine and residents collect their water from boreholes or local protected springs. Since 2004, periodic mass chemotherapy in schools has been conducted on three occasions as part of Child Health Days Plus, although the coverage has been variable, ranging from 30 to 65%; no community-based treatment has been implemented.

### Census and recruitment

Between June and August 2008, all households in the sub-county were visited for census enumeration, and a standard questionnaire was used to record the demographics of each household and information on asset ownership, crowding, water and sanitation and construction. Information on ownership of household assets was used to construct a wealth index for each household using principal component analysis (PCA), as described in [16]. The resulting score was divided into quintiles, to provide a categorical measure of relative socio-economic status. Household locations, health care facilities and major infrastructure were mapped using a hand-held eTrex global positioning system (Garmin Ltd., Olathe, KS). Geographic data were compiled and maps created using ArcGIS 9.2 (Environmental Systems Research Institute Inc., Redlands, CA, USA). Subsequently, four representative contiguous villages were purposively selected on the basis of size and accessibility, and all households in these villages re-visited between September and December 2008. Adults and parents/guardians of children had the purpose of the study explained to them in the language they felt most comfortable with and were

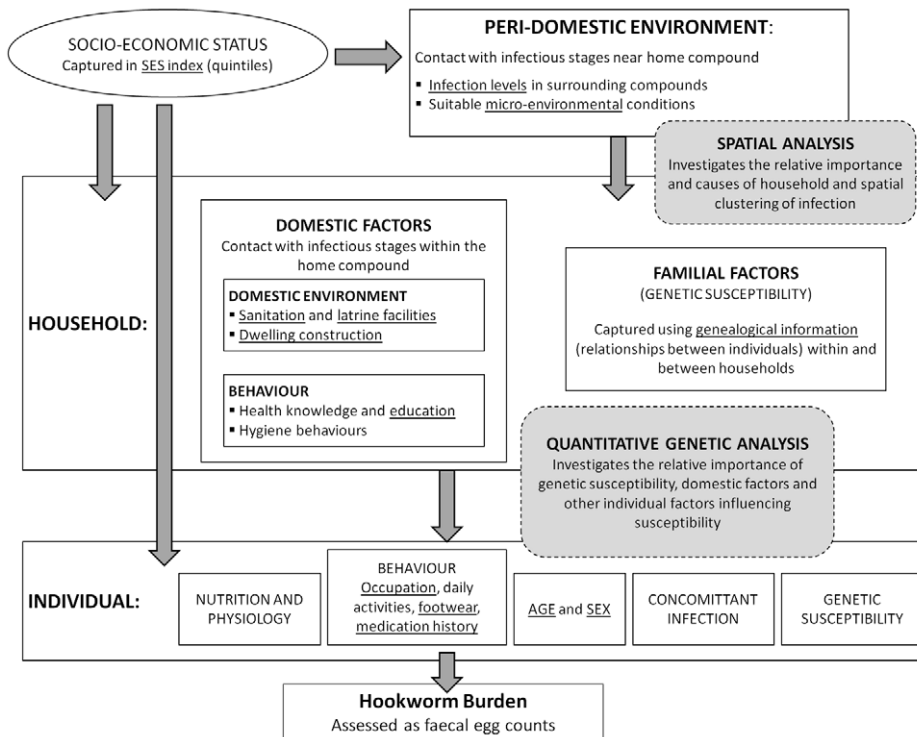
asked if they wished to participate. All willing residents aged over 6 months that were resident in the study site over the last 24 months, could be unambiguously tied to a single household, and did not attend school/work full time outside Mulanda parish, were eligible to participate. Excluded individuals were still offered parasitological examinations and treatment, but were not considered part of the data set for analysis. Signed informed consent was obtained from all adults and parents/guardians willing to participate and written assent from children aged 13–18 years. The study protocol was approved by the Makerere University Faculty of Medicine Research and Ethics Committee (#2008-043) Uganda National Council of Science and Technology (#HS 476) and London School of Hygiene and Tropical Medicine Ethics Committee (#5261).

### Procedures

Participants were asked to provide two daily consecutive stool samples (26.3% of participants provided only one sample) which were examined in duplicate by the Kato-Katz technique (using 41.7 mg slides) within 45 minutes of preparation and infection intensities expressed as eggs per gram (epg) of faeces using the arithmetic mean of all slides. Infected individuals were treated with a single 400 mg dose of albendazole. A standardized questionnaire was administered to adults and to primary carers of children to record details of previous anthelmintic treatment (user reported), education-level, school attendance and occupation and protective behaviours. Household heads were also interviewed about relationships between all household members, and about first- or second-degree relatives living in other households in the study villages, thereby identifying genetic links within and between households. Individuals were then defined as belonging to the same extended pedigree if they were related to anyone else in the pedigree or were married to anyone in the pedigree. Absent parents or family members connecting residents were assigned identification numbers and missing phenotype data. Pedigrees were assembled and indexed using PEDSYS [17] and visualised using Cranefoot [17], and any doubtful pedigree relationships were confirmed by re-interviewing selected household members.

### Analysis overview

The transmission dynamics of soil-transmitted helminth infection, including hookworm, are primarily determined by the number of worms present in the host (infection intensity), rather than the number of hosts infected [18]; analysis therefore focused on quantitative egg counts (intensity of infection), an indirect measure of worm burden. Data analysis was guided by a conceptual framework which takes account of the hierarchical relationships between peri-domestic, household and individual factors (Figure 1). In this framework, we utilise two complementary analytical approaches which both exploit correlation structures between individuals to evaluate aggregation in egg counts according to different sets of relationship, namely spatial and genetic relationships. Independent households within the same family compound shared the same location and peri-domiciliary environment and as such were treated as single units for both spatial and genetic analysis. Thus, in this report "household" refers to a family compound. Spatial analysis adopted a Bayesian geo-statistical approach incorporating both spatial correlation and non-spatial clustering at a household-level, whilst accounting for included covariates. Analysis of genetic factors was based on genetic variance component analysis whereby the relative importance of genetic susceptibility, other domestic factors and (non-genetic) individual factors in determining infection intensity was assessed. This was done via a frequentist variance partition



**Figure 1. Representation of the relationships between peri-domestic, household and individual variables and hookworm burden.** Includes the areas of this framework addressed by complementary analysis strategies used in this paper. For example, socio-economic factors (assessed, for example, by asset ownership) act through a number of inter-related proximate determinants, which include environmental factors (such as availability of water and sanitation), behavioural factors (such as health seeking behaviours) and nutrition (including anthropometric status and diet) which in turn may affect the risk of an individual being infected with hookworm. Covariates with available data are underlined. Both analyses address hookworm intensity (epg): the spatial analysis uses a negative binomial model, incorporating an over dispersion parameter ( $k$ ) to account for extra-Poisson variation; quantitative genetic analysis (genetic variance component analysis) uses a linear regression model, therefore data was log-transformed prior to analysis. These methods both exploit correlation structures in egg counts: for example, if there is a genetic basis to susceptibility, correlation due to genetic effects may be expected to decrease with distance in the pedigree, whereas if exposure to infective hookworm larvae most commonly occurs outside the household (for example in agricultural areas or defined defecation sites away from the household [42]), correlation due to environmental effects may decrease slowly with physical distance between households. doi:10.1371/journal.pntd.0000713.g001

approach, using degrees of relationship within and between households in order to partition variation into its genetic, domestic and other causes.

### Preliminary analysis

Egg counts are usually over-dispersed, and as such are often described well by the negative binomial distribution [19]. Negative binomial regression models were therefore applied to investigate factors associated with the intensity of hookworm infection. These models include an over-dispersion parameter,  $k > 0$ , which incorporates extra-Poisson variation. Preliminary investigation of covariates followed a frequentist approach, with standard errors adjusted for non-independence of individuals within households using robust Huber/White/Sandwich variance estimates, using Stata version 10 (Stata Corporation, College Station, Texas). Explanatory variables significant at the 10% significance level were entered into the multivariate negative binomial regression models, and backwards-stepwise elimination was used to generate a minimum adequate model; excluded covariates ( $P > 0.05$ ) were retested in the minimal model.

Quantitative genetic analysis, however, relies on linear regression, which assumes the outcome to be normally distributed. The data were therefore log-transformed, and multivariate linear regression models developed following the same stepwise procedures described above.

### Bayesian negative binomial spatial models

Spatial variation in infection intensity (and spatial clustering, if evident) was investigated by fitting a negative binomial distribution to the total egg count using the Bayesian spatial model of Alexander et al (2000), assuming a spatially constant  $k$  and including grams of faeces examined as an offset. Models were developed in WinBUGS Version 14. Between-household (non-spatial) variation was accounted for by a household-level random effect,  $u_j$ , with an exchangeable correlation matrix (mean 0, variance  $\sigma_1^2$ ). Fixed effect parameters from the negative binomial regression models are presented as density ratios (DR), that is,  $\exp(\beta)$  where  $\beta$  is the vector of regression coefficients. For ease of interpretation,  $100 \times (\text{DR} - 1)$  provides the estimated percentage change in egg counts for each unit increase in the independent variable. The spatial random effect,  $v_j$ , was modelled as a stationary Gaussian process with a mean 0, variance  $\sigma_2^2$  and correlation function  $\exp(-\alpha d_{kl})$ , where  $d_{kl}$  is the straight-line distance between households  $k$  and  $l$ . The geographical variability,  $\sigma_2^2$ , represents the spatially constructed component of variability on a log scale; the exponential of this parameter,  $\exp(v_j)$ , yields a ratio by which the house's mean egg count is higher or lower than expected, which can be thought of as a standardised parasite ratio (SPR) [11]. The smoothing parameter  $\alpha$  measures the rate at which the spatial correlation decays to zero with increasing distance;  $\ln(2\alpha)$  is the 'half-distance' (i.e. the distance over which

the correlation decays by half), and  $3\alpha$  is the distance at which the correlation reduces by 95%. Thus, the modelled mean egg counts and household clustering incorporate a degree of smoothing, which is dependent on the fitted spatial correlation.

Fixed regression coefficients were assigned diffuse priors functionally equivalent to a vague normal prior with mean 0 and large variance; spatial and non-spatial variance components ( $\sigma_1^2$  and  $\sigma_2^2$ ) were assigned gamma distributions; the over-dispersion parameter  $k$  was also assigned a gamma distribution with density proportional to  $k^{0.5} \exp(-0.01k)$  so that, *a priori*,  $k$  had a high variance; a weakly-informative uniform prior was assigned to  $\alpha$ , with parameters defined by semivariograms estimated on the basis of logarithmically transformed count data (i.e.  $\ln(\text{epg}+1)$ ), averaged by household, using the R module *GeoR*. We ran 35,000 iterations, with the first 10,000 discarded ('burn in') and sub-sampled every 5<sup>th</sup> observation, giving a final sample size of 5000 for which model parameters were estimated. Convergence was evaluated on the basis of inspection of sample traces, and inference was based on the better fitting model using the deviance information criteria (DIC) as a goodness of fit measure.

### Quantitative genetic analysis

Heritability estimation was undertaken using genetic variance component analysis implemented in SOLAR 4.2.0 (Southwest

Foundation for Biomedical Research, Texas) [20]. A series of mixed linear regression models were fitted to logarithmically-transformed arithmetic mean egg counts, i.e.  $\ln(\text{epg}+1)$ , using maximum likelihood procedures. Residual kurtosis for this trait was within acceptable range (i.e.  $<0.8$ ). Additive genetic effects were separated from household and other individual effects through incorporation of a kinship coefficient matrix, thus allowing the covariance between relatives to depend upon the degree of relatedness between them [21,22]. Chi-squared testing based on likelihood ratios was used to compare nested models and test the significance of variance parameters. Contributions of additive genetics (heritability,  $h^2$ ) and household ( $c^2$ ) were standardised by dividing by the total phenotypic variation. The relative contribution of fixed covariates to total phenotypic variance was estimated by comparing the trait standard deviation in models with and without fixed covariates.

### Results

Complete questionnaire and parasitological data were available for 1,803 individuals (76.5% total population) and pedigree information for 1,687 (71.5%) individuals. There were 341 compounds, with 1–19 (median 5) phenotyped individuals per compound; an average compound comprised 2.9 phenotyped children (0–15 years). Comparison of census data for those that

**Table 1.** Baseline characteristics of study participants by demographic group.

Personal characteristics:	NUMBER (%) STRATIFIED BY DEMOGRAPHIC GROUP		
	<5 years; n = 385	5–15 years; n = 659	≥16 years; n = 831
Hookworm infection:			
Prevalence, % (n/total)	26.5%	33.2%	51.5%
Geometric mean egg count, epg	82	87	196
Sex: male	187 (48.6%)	323 (51.3%)	300 (38.1%)
Anthelmintic in previous 6 months	137 (35.9%)	423 (67.5%)	141 (17.9%)
Wears shoes outside home	110 (28.6%)	101 (15.3%)	457 (58.0%)
Primary carer with any education	285 (74.4%)	437 (69.8%)	-
Currently attends school	-	537 (86.3%)	-
Ever attended school	-	-	432 (54.8%)
Occupation:			
Farmer	-	-	324 (41.1%)
Formal employment	-	-	58 (7.4%)
Student	-	-	55 (7.0%)
None	-	-	330 (41.9%)
<b>Household characteristics: NUMBER (%) n = 435</b>			
Toilet facilities:			
Covered pit/VIP <sup>a</sup>	88 (20.2%)		
Uncovered pit	326 (74.9%)		
None	21 (4.8%)		
Floor type:			
Mud/beaten earth	392 (90.1%)		
Cement	43 (9.9%)		
Member with formal income	63 (14.5%)		
Educated household head <sup>b</sup>	29 (6.6%)		

<sup>a</sup>VIP; ventilated improved pit latrine.

<sup>b</sup>Household head with education above secondary (higher education or vocational training).

doi:10.1371/journal.pntd.0000713.t001

participated and those that chose not to participate suggests that our final study sample under-sampled adult males ( $p < 0.001$ ). There were however no statistical difference in the size or relative socio-economic status of participant and non-participant households [23]. Baseline characteristics of the study population are shown in Table 1.

The prevalence of hookworm was 39.3%, with the majority of infections (87.7%)  $\leq 1000$  epg. The geometric mean intensity of infection was 134 epg (95% confidence intervals [CI]: 118–153). Other helminth species were rare: prevalence of *Schistosoma mansoni* was 1.7%, *Trichuris trichiura* 0.2% and *Ascaris lumbricoides* 0.2%. The prevalence of hookworm did not differ significantly by sex. Prevalence rose steadily with age (Figure 2A), whilst intensity increased substantially from  $\sim 35$  years. Mean intensity among over 60 year olds was 2,318 epg (95% CI 864 to 3,773) in females and 936 (95% CI 387 to 1,485) in males (Figure 2B). However, much of this increase is due to very high egg counts in a few older individuals, and the increase in median intensity (in those infected) among the elderly was less marked (Figure 2C).

### Risk factors

Table 2 presents the results of the spatial model for hookworm intensity. In multivariate analysis there was no evidence of a difference between males and females, although intensity did rise with age. As expected, intensity was associated with factors influencing environmental exposure to infection, as well as socio-economic indicators and treatment history. Specifically, intensity was higher among older individuals and individuals who reported not receiving anthelmintic treatment in the previous six months, walking barefoot outside the home, with living in a household with a mud floor and with low education of household head. Intensity was also higher among individuals who provided two stool samples, a likely reflection of the increased likelihood of detecting an infection in this group.

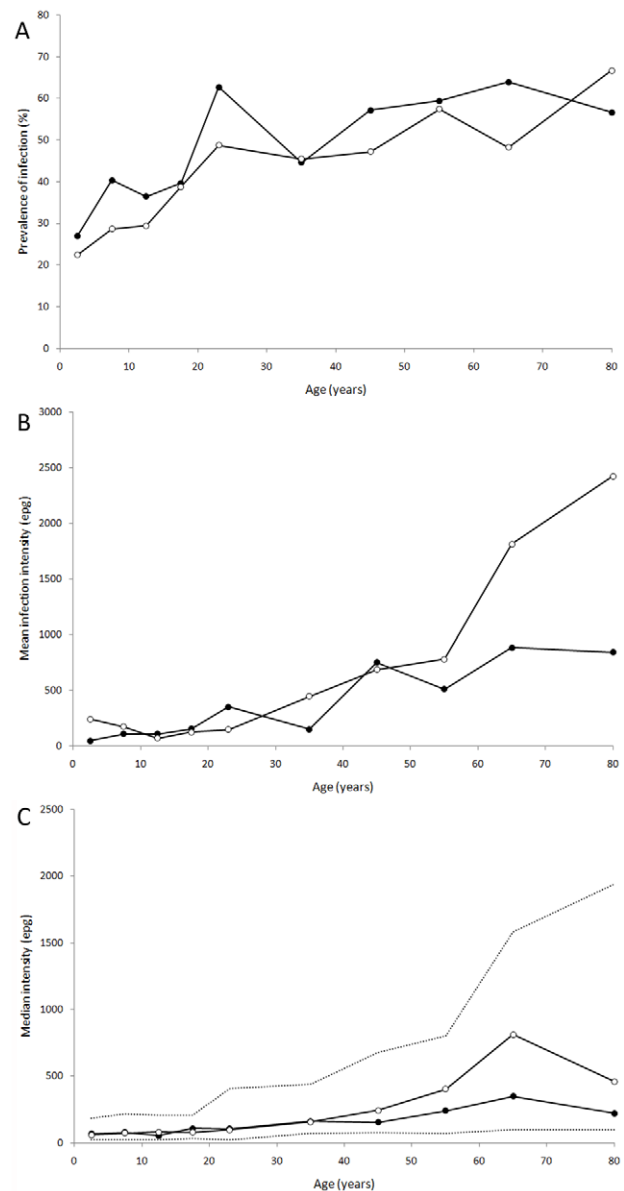
### Residual spatial clustering

Intensity of infection exhibited significant household and spatial clustering, as indicated by the improved fit of models including household-level spatial and non-spatial random effects. Of the total variation between households, 81% could be explained by small-scale spatial clustering and 19% by non-spatial household clustering. The range of spatial correlation was estimated to be 82 m and was reduced by a half over a distance of 19 m. Figure 3 shows the model-based estimates of spatial variation in infection intensity and highlights three separate areas of high infection intensity.

### Household and genetic determinants of infection

Genetic variance component models were used to define the contribution of hereditary and environmental influences on intensity of infection, thereby providing an estimate of the total cumulative effect of individual genes (the “additive genetic” effect, or heritability) on overall variation. Of the 1,687 people for whom relationship information was available, 1,624 could be grouped into 155 pedigrees, with two large multi-generational pedigree comprising 690 and 173 phenotyped individuals respectively, 21 smaller pedigrees of 10–60 phenotyped individuals and 69 nuclear families ( $< 10$  members); 63 participants had no phenotyped relatives in the study area. When considering up to 8 degrees of relatedness (over 5 generations), data was available for 11,509 informative relative pairs; this includes 2,627 first degree relatives (e.g. siblings), 1,968 second degree relatives (e.g. grandparent/grandchild) and 2,291 third degree relatives (e.g. cousins).

Variance parameter estimates and corresponding model log-likelihoods are provided in Table 3. Intensity was best described by a saturated model which included both additive genetic and shared household effects. As shown in Table 3, risk factors identified as significantly associated with log-transformed infection intensity are consistent with results from the above spatial negative binomial model. When adjusting for these covariates (which explained 18% of total variation in log-transformed egg counts), heritability was estimated as 11.2% (Standard error [SE] 5.7%), with a further 17.8% of residual variation (SE 3.7%) attributable to other unmeasured domestic factors; 71% of residual variation (58% of total variation) remained unexplained by all factors in this model.



**Figure 2. Age-sex distributions for hookworm infection.** (A) prevalence of hookworm infection, (B) mean intensity of infection, in epg, and (C) median intensity of infection (in epg) for those infected only (dashed lines indicate the population interquartile range). Males, closed circles; Females, open circles. doi:10.1371/journal.pntd.0000713.g002

**Table 2.** Final estimates for covariates and variance parameters for hookworm infection including variance parameter estimates.

COVARIATES	(i) SPATIAL NEGATIVE BINOMIAL MODEL		(ii) GENETIC VCA LINEAR REGRESSION MODEL	
	Estimate	(95% BCI)	Estimate	(95% CI)
INDIVIDUAL CHARACTERISTICS:				
Age (per year)	1.08	(1.08,1.12)	1.04	(1.03,1.05)
Two stool samples (vs. one)	1.45	(0.83,2.27)	1.62	(1.23,2.12)
Barefoot (vs. shoes or sandals)	2.86	(1.71,4.51)	1.43	(1.12,1.84)
Anthelmintic within 6 months	0.29	(0.17,0.46)	0.51	(0.39,0.66)
SOCIO-ECONOMIC INDICATORS:				
Educated primary carer <sup>1</sup>	0.14	(0.04,0.37)	0.61	(0.44,0.84)
Educated household head <sup>2</sup>	-		0.66	(0.31,1.22)
Formal income	-		0.67	(0.44,1.03)
HOUSEHOLD CHARACTERISTICS:				
Beaten earth flooring	20.99	(7.63,47.71)	3.03	(1.72,5.37)
>500m from rocky area	-		1.58	(1.14,2.20)
RANDOM EFFECT PARAMETERS (SPATIAL NEGATIVE BINOMIAL MODEL):				
$k$ (over-dispersion parameter)	0.07	(0.0005)		
$\sigma_1^2$ (non-spatial household variation)	0.31	(0.04,1.05)		
$\sigma_2^2$ (spatial household variation)	1.31	(0.03,2.67)		
$\alpha$ (smoothing parameter)	27.23	(5.6,173.4)		
VARIANCE COMPONENTS AS PROPORTIONS OF RESIDUAL VARIATION (SE) (GENETIC VCA MODEL):				
Additive genetic			0.11	(0.06)
Shared domestic environment			0.18	(0.04)
Individual-specific			0.71	(0.05)

<sup>1</sup>Household head with education above secondary.

<sup>2</sup>Primary carer with any level of education (primary incomplete and above).

(i) Bayesian negative binomial geospatial model used to estimate the scale and importance of residual spatial correlation in mean household prevalence. The estimate of the negative binomial regression covariate parameters represent density ratio (DR) of egg counts. (ii) Genetic variance component analysis used to estimate the proportion of residual variance attributable to additive genetics, individual (non-genetic) and shared household effects. The estimate of the genetic VCA covariate parameters have been transformed back to the original egg count scale (exponentiated regression coefficients), and are thus multiplicative and directly comparable with the negative binomial DRs. It is not possible to standardise variance parameters in the spatial negative binomial model.

The estimate of the negative binomial regression coefficient parameters represent the density ratio (DR) of egg counts, and that of the genetic VCA linear regression model the exponentiated regression coefficients. Both are multiplicative.

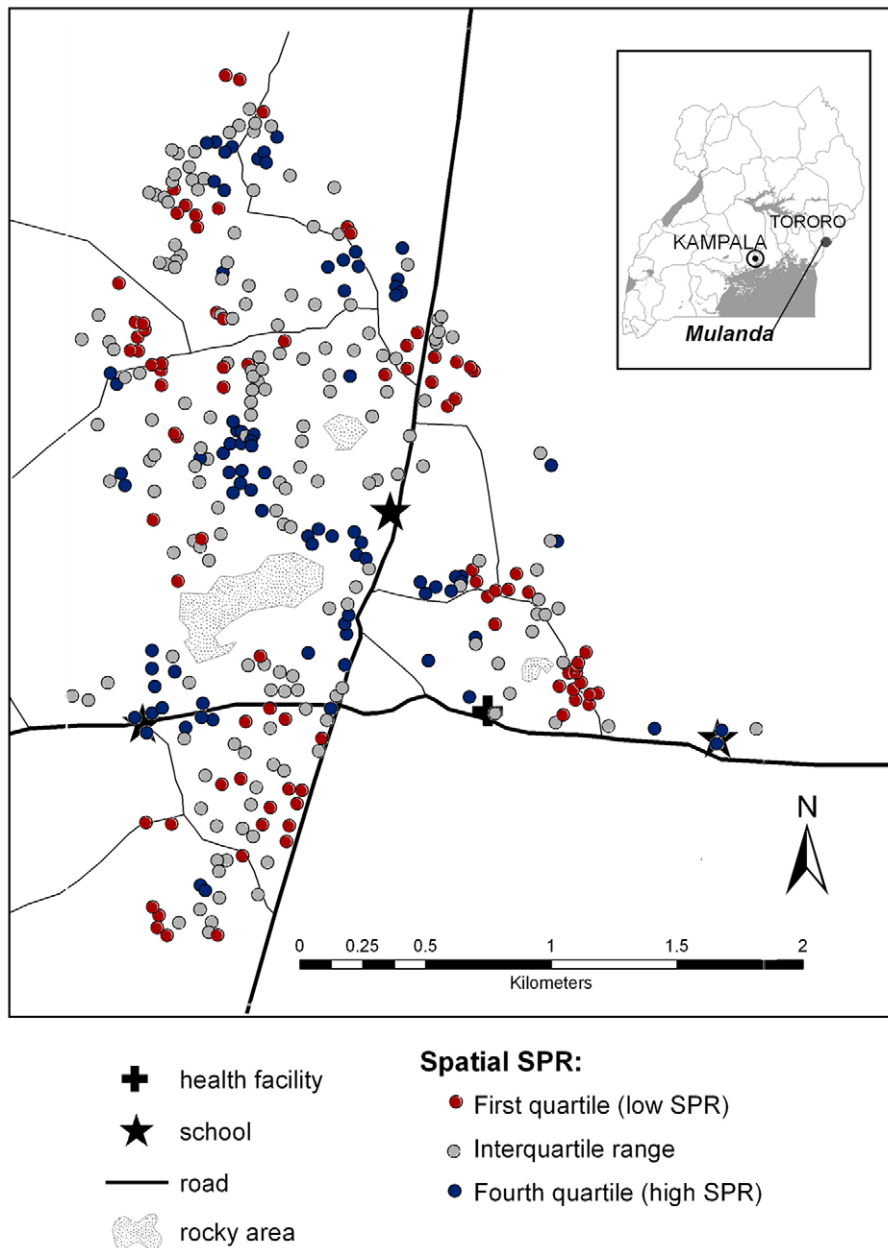
doi:10.1371/journal.pntd.0000713.t002

## Discussion

Marked heterogeneity within communities has been long recognised for intestinal nematode infections [1,10–12,15,18,24–26], although there have been few studies undertaken in post-treatment settings in sub-Saharan Africa. This analysis sought to comprehensively assess major sources of heterogeneity (spatial, genetic and household) in the intensity of hookworm infection in a Ugandan population using both spatial and quantitative genetic modelling approaches. In keeping with previous studies [27–30] the highest infection intensities were found among older residents, and we identified a number of individual and household risk factors. After taking these features into account, we revealed substantial spatial clustering of infection intensity in this community, with the relatively small spatial range (81m) suggesting that much of the exposure to infection is concentrated in the peri-domiciliary environment. Focusing further on the household/family level, we estimate that 11% of residual variability in egg counts could be attributed to genetic differences between individuals (heritability) after accounting for individual and household risk factors. Unmeasured factors associated with the domestic environment explained a further 18% of residual variation. Results from quantitative genetic analysis also demon-

strate the central influence of the domestic environment, with household-level covariates and shared household accounting for nearly one fifth of total variation in hookworm intensity.

The spatial heterogeneity of parasitic infections has gained increased prominence in recent years [31,32]. Spatial patterns may not always be evident in maps of unadjusted mean infection intensity because of sampling variation caused by, for example, socio-demographic factors and aggregation of parasite egg counts [33]. The methods adopted here to assess spatial structure have the particular advantage of taking into account the highly skewed distribution of parasite counts, whilst adjusting for covariates [34,35]. Similar ranges of hookworm spatial clustering have also been reported in Brazil [11] and are consistent with household studies of urinary schistosomiasis [36] and malaria [37]. In contrast, studies of hookworm infection in Cote d'Ivoire [38,39] and Kenya [40] saw no evidence of household- or school-level spatial clustering at this scale. Differences between sites may be explained by the relative importance of domestic, peri-domestic and public domains in defining contact with infectious stages [41]. For example, previous studies have suggested exposure to infective hookworm larvae commonly occurs outside the household (for example in agricultural areas, schools or defined defecation sites [42]), although our results suggest that in this community



**Figure 3. Map of the small-scale spatial heterogeneity of hookworm intensity in Mulanda, after adjusting for individual and household risk factors.** Each dot represents a compound, with the shading showing the quartile of the standardised parasite ratio (SPR; ratio by which the compound's mean log egg count is higher or lower than expected) as derived by the Bayesian spatial model. High infection intensities are seen in the south-west, north-west and east of the study site, whilst low intensity infections are seen in the central region. Inset: location of the study site in eastern Uganda.

doi:10.1371/journal.pntd.0000713.g003

transmission occurs within or close to households, consistent with the significant household clustering observed. Residual spatial clustering of hookworm infection in this community may be attributable to small-scale variation in unmeasured environmental factors, such as soil type or vegetation cover [10,43].

Previous quantitative genetic analysis of helminth infection within human populations suggests roles for both extrinsic (domestic environmental) and intrinsic (genetic) factors in clustering of intestinal nematode infections within households [12,14,24,25,44,45]. To date, the only study to investigate heritability in an African setting, conducted in Zimbabwe, suggested heritability of hookworm infection intensity was 0.37,

although this study failed to account for household effects, limiting interpretation of findings. Our analysis suggested that, after accounting for household effects and covariates, heritability in this Ugandan population is relatively modest at 0.11, comparable to findings from Papua New Guinea [12] but lower those reported by studies conducted in high transmission settings in Brazil (0.20–0.25) (*Bethony, unpublished data*) [13]. Heritability is a population-specific measure, and as such direct comparison of genetic contributions between sites can be confounded by both the differing importance of environmental factors and by variation in genetic polymorphisms of the pathogen/host [46,47]. Contributions of genetic and household effects may also be influenced by

**Table 3.** Variance component analysis of heritability ( $h^2$ ) and household effects ( $c^2$ ) of intensity of hookworm infection.

		VARIANCE COMPONENT ESTIMATES					
		Hookworm $\ln(\text{egp}+1)$					
MODELS		LnL	$h^2$	$c^2$	$e^2$	LR test	P
<b>NO COVARIATES</b>							
1	Sporadic ( $e^2$ )	-2444			1		
2	Polygenic ( $h^2 + e^2$ )	-2367	0.380		0.620	A (2 vs. 1)	<0.001
3	Household ( $c^2 + e^2$ )	-2387		0.275	0.725	B (3 vs. 1)	<0.001
4	Saturated ( $h^2 + c^2 + e^2$ )	-2364	0.106	0.235	0.659	C (4 vs. 3)	0.02
						D (4 vs. 2)	<0.001
<b>WITH COVARIATES<sup>†</sup></b>							
1	Sporadic ( $e^2$ )	-2251			1		
2	Polygenic ( $h^2 + e^2$ )	-2196	0.337		0.663	A (2 vs. 1)	<0.001
3	Household ( $c^2 + e^2$ )	-2207		0.221	0.779	B (3 vs. 1)	<0.001
4	Saturated ( $h^2 + c^2 + e^2$ )	-2194	0.112	0.178	0.710	C (4 vs. 3)	0.02
						D (4 vs. 2)	<0.001

<sup>†</sup>Model adjusted for age (linear term), medication history, number of stool samples examined, footwear, household income, education level and residential location. Chi-squared test of nested models; A:  $h^2 \neq 0$  (not controlling for  $c^2$ ); B:  $c^2 \neq 0$  (not controlling for  $h^2$ ); C:  $h^2 \neq 0$  (after controlling for  $c^2$ ); D:  $c^2 \neq 0$  (after controlling for  $h^2$ ). doi:10.1371/journal.pntd.0000713.t003

sample size and pedigree and household structures, although studies do suggest that the impact of pedigree structure on precision and accuracy of heritability estimates may be minimal for purely additive genetic effects [48]. It is conceivable however that host genes may play a more important role in higher transmission settings. Alternatively, different immune responses, and thus genes, may influence initial versus re-infection, intensity [49,50] resulting in lower heritability estimates in this post-treatment setting. Alternatively, since previous treatment could not be measured accurately and instead relied upon participants reporting use of anthelmintic medication, this may have reduced heritability by adding to unexplained variation.

Individual and household risk factors for hookworm infection have been explored extensively, and our findings are in keeping with the literature (see review by Brooker et al (2004)[51], also [52–57]). Reassuringly, whilst the two analysis strategies looked at different outcomes (untransformed and log-transformed count data) the significance and magnitude of associations were very similar for each approach, confirming the validity of the findings. Although studies frequently observe higher burdens of hookworm in males than females [1,6,51,58], we could find no evidence of a sex difference here, suggesting an absence of sex-related differences in exposure for this population. Absence of latrine facilities was not associated with hookworm infection risk in this community, although this may be indicative of the relatively uniform distribution of poor quality uncovered pit latrines throughout the community [52,53]. Similarly, whilst relative socioeconomic group *per se* was not associated with hookworm infection, other indirect measures of household socio-economic status (household member with a formal income, education level of primary carer and household head, walking barefoot when outside the home) were risk factors. Socio-economic status is likely to influence exposure to hookworm infection risk via a number of mechanisms, which may include poor hygienic behaviour and health knowledge [59–62].

It is worth noting that for practical reasons we rely on egg counts as an indirect measure of worm burden, and whilst the relationship between egg production and worm burden in

hookworm infection is approximately linear, faecal egg counts are known to fluctuate daily, limiting Kato-Katz sensitivity and increasing measurement error [63]. This is reflected by our observation that mean infection intensities were higher for participants who provided two stool samples, suggesting that infections were missed in those that brought only one. As such, we may have underestimated sources of heterogeneity. Furthermore, biases inherent in using faecal egg counts as proxy for worm burden may have added to unexplained heterogeneity. However, the only real alternative for quantifying hookworm infection levels is actual worm burden, which is unfeasible to collect for large population-based studies. A further limitation is the inability to incorporate both genetic and spatial clustering into one model, and as such - as relatives are more likely to live close together - this analysis does risk over-estimating the role of each determinant.

In conclusion, this report provides one of the most comprehensive population-level epidemiological analyses of hookworm infection for sub-Saharan Africa. We demonstrate that, despite several rounds of school-based deworming, prevalence and intensity of infection amongst adults remains high in this community. Such observations have potentially important implications for mass deworming programmes, which currently target only school-aged children. There are currently few data quantifying the benefits of deworming in non-pregnant adult populations, but results presented here and elsewhere [28,64,65] suggest that extension of deworming efforts to adult populations may prove a valuable intervention. The feasibility of such an approach is demonstrated by ongoing lymphatic filariasis control, which includes co-administration of ivermectin and albendazole to all (non-pregnant) individuals aged above 4 years and represent the largest community-based helminth control programmes in Africa [66,67]. From a scientific perspective, the use of complementary spatial and genetic epidemiological approaches has allowed us to update our understanding of the ecology of infection, and has emphasised the continuing importance of household and peri-domestic factors, as well as host genetic relatedness, in influencing infection patterns even after the implementation of control measures.



## Supporting Information

### Checklist S1 STROBE checklist.

Found at: doi:10.1371/journal.pntd.0000713.s001 (0.09 MB DOC)

## Acknowledgments

We thank the residents of Mulanda who kindly participated in the study, and thank all members of the study team in Uganda for their excellent

## References

- Behnke JM, De Clercq D, Sacko M, Gilbert FS, Ouattara DB, et al. (2000) The epidemiology of human hookworm infections in the southern region of Mali. *Trop Med Int Health* 5: 343–354.
- Palmer DR, Bundy DAP (1995) Epidemiology of human hookworm and *Ascaris lumbricoides* infestations in rural Gambia. *East Afr Med J* 72: 527–530.
- Udonsi JK (1984) Studies on the co-occurrence of two species of human hookworm in a riverine community in Nigeria. *Tropenmed Parasitol* 35: 37–40.
- Udonsi JK (1984) *Necator americanus* infection: a cross-sectional study of a rural community in relation to some clinical symptoms. *Ann Trop Med Parasitol* 78: 443–444.
- Chandiwana SK, Bradley M, Chombo F (1989) Hookworm and roundworm infections in farm-worker communities in the large-scale agricultural sector in Zimbabwe. *J Trop Med Hyg* 92: 338–344.
- Asaolu SO, Holland CV, Jegede JO, Fraser NR, Stoddard RC, et al. (1992) The prevalence and intensity of soil-transmitted helminthiasis in rural communities in southern Nigeria. *Ann Trop Med Parasitol* 86: 279–287.
- Bradley M, Chandiwana SK, Bundy DA, Medley GF (1992) The epidemiology and population biology of *Necator americanus* infection in a rural community in Zimbabwe. *Trans R Soc Trop Med Hyg* 86: 73–76.
- Diggle P, Moyeed R, Tawn J (1998) Model-based geostatistics (with discussion). *Applied Statistics* 47: 299–350.
- Pfeiffer DU, Hugh-Jones M (2002) Geographical information systems as a tool in epidemiological assessment and wildlife disease management. *Rev Sci Tech* 21: 91–102.
- Saathoff E, Olsen A, Sharp B, Kvalsvig JD, Appleton CC, et al. (2005) Ecologic covariates of hookworm infection and reinfection in rural KwaZulu-natal/south Africa: a geographic information system-based study. *Am J Trop Med Hyg* 72: 384–391.
- Brooker S, Alexander N, Geiger S, Moyeed RA, Stander J, et al. (2006) Contrasting patterns in the small-scale heterogeneity of human helminth infections in urban and rural environments in Brazil. *Int J Parasitol* 36: 1143–1151.
- Breitling LP, Wilson AJ, Raiko A, Lagog M, Siba M, et al. (2008) Heritability of human hookworm infection in Papua New Guinea. *Parasitology* 135: 1407–1415.
- Quinnell RJ, Pullan RL, Brooker S, Geiger S, Correa-Oliveira R, et al. (2009) Genetic and household determinants of predisposition to human hookworm infection in a Brazilian community. Submitted.
- Pullan RL, Bethony JM, Geiger SM, Correa-Oliveira R, Brooker S, et al. (2009) Human helminth co-infection: no evidence of common genetic control of hookworm and *Schistosoma mansoni* infection intensity in a Brazilian community. *Int J Parasitol*.
- Williams-Blangero S, Blangero J, Bradley M (1997) Quantitative genetic analysis of susceptibility to hookworm infection in a population from rural Zimbabwe. *Hum Biol* 69: 201–208.
- Makinen VP, Karkkonen M, Wessman M, Groop P-H, Kanninen T, et al. (2005) High throughput pedigree drawing. *Eur J Hum Genet* 13.
- Dyke B (1999) PEDSYS: A Pedigree Data Management System. Southwest Foundation for Biomedical Research., San Antonio, TX.
- Anderson RM, May RM (1985) Helminth infections of humans: mathematical models, population dynamics, and control. *Advances in Parasitology* 24: 1–101.
- Anderson RM (1993) Epidemiology. In: Cox FEG, ed. *Modern Parasitology*. Oxford: Blackwell Science. pp 75–116.
- Almasy L, Blangero J (1998) Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 62: 1198–1211.
- Khoury MJ, Beaty TH, Cohen BH (1993) *Fundamentals of genetic epidemiology*. New York: Oxford University Press.
- Rice TK, Borecki IB (2001) Familial resemblance and heritability. *Adv Genet* 42: 35–44.
- Pullan RL, Bukirwa H, Staedke SG, Snow RW, Brooker S (2010) *Plasmodium* infection and its risk factors in eastern Uganda. *Malar J*: In press.
- Chan L, Bundy DA, Kan SP (1994) Genetic relatedness as a determinant of predisposition to *Ascaris lumbricoides* and *Trichuris trichiura* infection. *Parasitology* 108(Pt 1): 77–80.
- Chan L, Bundy DA, Kan SP (1994) Aggregation and predisposition to *Ascaris lumbricoides* and *Trichuris trichiura* at the familial level. *Trans R Soc Trop Med Hyg* 88: 46–48.
- Shapiro AE, Tukahebwa EM, Kasten J, Clarke SE, Magnussen P, et al. (2005) Epidemiology of helminth infections and their relationship to clinical malaria in southwest Uganda. *Trans R Soc Trop Med Hyg* 99: 18–24.
- Bethony J, Chen J, Lin S, Xiao S, Zhan B, et al. (2002) Emerging patterns of hookworm infection: influence of aging on the intensity of *Necator* infection in Hainan Province, People's Republic of China. *Clin Infect Dis* 35: 1336–1344.
- Brooker S, Jardim-Botelho A, Quinnell RJ, Geiger SM, Caldas IR, et al. (2007) Age-related changes in hookworm infection, anaemia and iron deficiency in an area of high *Necator americanus* hookworm transmission in south-eastern Brazil. *Trans R Soc Trop Med Hyg* 101: 146–154.
- Needham C, Kim HT, Hoa NV, Cong LD, Michael E, et al. (1998) Epidemiology of soil-transmitted nematode infections in Ha Nam Province, Vietnam. *Tropical Medicine & International Health* 3: 904–912.
- Ye XP, Wu ZX, Sun FH (1994) The population biology and control of *Necator americanus* in a village community in south-eastern China. *Ann Trop Med Parasitol* 88: 635–643.
- Hay SI, Randolph SE, Rogers DJ (2000) *Remote Sensing and Geographical Information Systems in Epidemiology*. London, UK: Academic Press.
- Pfeiffer DU, Robinson TP, Stevenson M, Stevens KB, Rogers DJ, et al. (2008) *Spatial analysis in epidemiology*. Oxford, UK: Oxford University Press.
- Bernardinelli L, Montomoli C (1992) Empirical Bayes versus fully Bayesian analysis of geographical variation in disease risk. *Stat Med* 11: 983–1007.
- Bliss CA, Fisher RA (1953) Fitting the negative binomial to biological data and a note on the efficient fitting of the negative binomial. *Biometrics* 9: 176–200.
- Alexander N (2000) Spatial modelling of individual-level parasite counts using the negative binomial distribution. *Biostatistics* 1: 453–463.
- Clennon JA, King CH, Muchiri EM, Karuiki HC, Ouma JH, et al. (2004) Spatial patterns of urinary schistosomiasis infection in a highly endemic area of coastal Kenya. *Am J Trop Med Hyg* 70: 443–448.
- Brooker S, Clarke S, Njagi JK, Polack S, Mugo B, et al. (2004) Spatial clustering of malaria and associated risk factors during an epidemic in a highland area of western Kenya. *Trop Med Int Health* 9: 757–766.
- Utzinger J, Muller I, Vounatsou P, Singer BH, N'Goran EK, et al. (2003) Random spatial distribution of *Schistosoma mansoni* and hookworm infections among school children within a single village. *J Parasitol* 89: 686–692.
- Matthys B, Tschannen AB, Tian-Bi NT, Comoe H, Diabate S, et al. (2007) Risk factors for *Schistosoma mansoni* and hookworm in urban farming communities in western Cote d'Ivoire. *Tropical Medicine & International Health* 12: 709–723.
- Handzel T, Karanja DM, Addiss DG, Hightower AW, Rosen DH, et al. (2003) Geographic distribution of schistosomiasis and soil-transmitted helminths in Western Kenya: implications for anthelmintic mass treatment. *Am J Trop Med Hyg* 69: 318–323.
- Cairncross S, Blumenthal U, Kolsky P, Moraes L, Tayeh A (1996) The public and domestic domains in the transmission of disease. *Trop Med Int Health* 1: 27–34.
- Schad GA, Nawalinski TA, Kochar V (1983) *Human ecology and the distribution and abundance of hookworm populations. Human ecology and infectious diseases*: Academic Press.
- Pullan R, Bethony J, Geiger S, Cundill B, Correa-Oliveira R, et al. (2008) Human helminth co-infection: an analysis of spatial heterogeneity and household and environmental risk factors in a Brazilian community. *PLoS Negl Trop Dis* 2: e352.
- Williams-Blangero S, Subedi J, Upadhyay RP, Manral DB, Rai DR, et al. (1999) Genetic analysis of susceptibility to infection with *Ascaris lumbricoides*. *Am J Trop Med Hyg* 60: 921–926.
- Williams-Blangero S, McGarvey ST, Subedi J, Wiest PM, Upadhyay RP, et al. (2002) Genetic component to susceptibility to *Trichuris trichiura*: evidence from two Asian populations. *Genet Epidemiol* 22: 254–264.
- Yen-Revollo JL, Van Booven DJ, Peters EJ, Hoskins JM, Engen RM, et al. (2009) Influence of ethnicity on pharmacogenetic variation in the Ghanaian population. *Pharmacogenomics J*.
- Hill AV, Yates SN, Allsopp CE, Gupta S, Gilbert SC, et al. (1994) Human leukocyte antigens and natural selection by malaria. *Philos Trans R Soc London Ser B* 346: 379–385.
- Ekstrom CT (2009) The impact of pedigree structure on heritability estimates. *Human Heredity* 68: 243–251.
- Quinnell RJ, Woolhouse ME, Walsh EA, Pritchard DI (1995) Immunoepidemiology of human necatoriasis: correlations between antibody responses and parasite burdens. *Parasite Immunol* 17: 313–318.

work, in particular Annet Enzaru and Betty Kagoro from the Vector Control Division.

## Author Contributions

Conceived and designed the experiments: RLP RJQ SB. Performed the experiments: RLP. Analyzed the data: RLP. Contributed reagents/materials/analysis tools: NBK SB. Wrote the paper: RLP RJQ SB. Provided guidance and technical expertise: NBK.

50. Quinnell RJ, Pritchard DI, Raiko A, Brown AP, Shaw MA (2004) Immune responses in human necatoriasis: association between interleukin-5 responses and resistance to reinfection. *J Infect Dis* 190: 430–438.
51. Brooker S, Bethony J, Hotez PJ (2004) Human hookworm infection in the 21st century. *Adv Parasitol* 58: 197–288.
52. Sorensen RE, Ismail MM, Amarasinghe DK, Hettiarachchi I, Dassenaieke TS (1994) The effect of the availability of latrines on soil-transmitted nematode infections in the plantation sector in Sri Lanka. *Am J Trop Med Hyg* 51: 36–39.
53. Chongsuvivatwong V, Pas-Ong S, McNeil D, Geater A, Duerawee M (1996) Predictors for the risk of hookworm infection: experience from endemic villages in southern Thailand. *Trans R Soc Trop Med Hyg* 90: 630–633.
54. Holland CV, Taren DL, Crompton DW, Nesheim MC, Sanjur D, et al. (1988) Intestinal helminthiasis in relation to the socioeconomic environment of Panamanian children. *Soc Sci Med* 26: 209–213.
55. Bethony J, Williams JT, Kloos H, Blangero J, Alves-Fraga L, et al. (2001) Exposure to *Schistosoma mansoni* infection in a rural area in Brazil. II: household risk factors. *Trop Med Int Health* 6: 136–145.
56. Raso G, Utzinger J, Silue KD, Ouattara A, Yapi A, et al. (2005) Disparities in parasitic infections, perceived ill health and access to health care among poorer and less poor schoolchildren of rural Côte d'Ivoire. *Trop Med Int Health* 10: 42–57.
57. Hotez P (2007) Hookworm and poverty. *Ann N Y Acad Sci* 1136: 38–44.
58. Bundy DA (1988) Gender-dependent patterns of infections and disease. *Parasitol Today* 4: 186–189.
59. Kamunvi F, Ferguson AG (1993) Knowledge, attitudes and practices (KAP) of human intestinal helminths (worms) in two rural communities in Nyanza Province, Western Kenya. *East Afr Med J* 70: 482–490.
60. de Silva N, Jayapandi VP, de Silva HJ (1996) Socioeconomic and behavioral factors affecting the prevalence of geohelminths in preschool children. *Southeast Asian J Trop Med Public Health* 27: 36–42.
61. Phiri K, Whitty CJ, Graham SM, Ssembatya-Lule G (2000) Urban/rural differences in prevalence and risk factors for intestinal helminth infection in southern Malawi. *Ann Trop Med Parasitol* 94: 381–387.
62. Traub RJ, Robertson ID, Irwin P, Mencke N, Andrew-Thompson RC (2004) The prevalence, intensity and risk factors associated with geohelminth infection in tea-growing communities of Assam, India. *Tropical Medicine and International Health* 9: 668–701.
63. Anderson RM, Schad GA (1985) Hookworm burdens and faecal egg counts: an analysis of the biological basis of variation. *Trans R Soc Trop Med Hyg* 79: 812–825.
64. Olsen A, Magnussen P, Ouma JH, Andreassen J, Friis H (1998) The contribution of hookworm and other parasitic infections to haemoglobin and iron status among children and adults in western Kenya. *Trans R Soc Trop Med Hyg* 92: 643–649.
65. Gilgen D, Mascie-Taylor CG (2001) The effect of anthelmintic treatment on helminth infection and anaemia. *Parasitology* 122 Pt 1: 105–110.
66. WHO (2002) Preparing and implementing a national plan to eliminate Lymphatic Filariasis. Geneva: World Health Organisation.
67. WHO (2009) Global programme to eliminate lymphatic filariasis – Progress report on mass drug administration in 2008. *Weekly Epidemiological Record* 84: 437–444.