GENETIC ANALYSIS OF SUSCEPTIBILITY TO INFECTION WITH ASCARIS LUMBRICOIDES

SARAH WILLIAMS-BLANGERO, JANARDAN SUBEDI, RAM PRASAD UPADHAY, DEEPAK B. MANRAL, DEV R. RAI, BHARAT JHA, EDWARD S. ROBINSON, AND JOHN BLANGERO

Abstract. Epidemiologic studies of helminthic infections have shown that susceptibility to these parasites frequently aggregates in families, suggesting the possible involvement of genetic factors. This paper presents a genetic epidemiologic analysis of Ascaris lumbricoides infection in the Jirel population of eastern Nepal. A total of 1,261 individuals belonging to a single pedigree were assessed for intensity of Ascaris infection at two time points. Following an initial assessment in which all individuals were treated with albendazole, a follow-up examination was performed one year later to evaluate reinfection patterns. Three measures of worm burden were analyzed, including eggs per gram of feces, direct worm counts, and worm biomass (weight). For all traits, variance component analysis of the familial data provided unequivocal evidence for a strong genetic component accounting for between 30% and 50% of the variation in worm burden. Shared environmental (i.e., common household) effects account for between 3% and 13% of the total phenotypic variance.

Infection with roundworm (Ascaris lumbricoides) remains a major international public health concern. 1,2 Roundworm is the most prevalent of the intestinal helminthic infections, affecting one-fourth of the world’s population. 3 Severe cases can cause serious morbidity and mortality due to intestinal blockage. 4-6 Chronic ascariasis has been implicated in the development and persistence of malnutrition in children, and may have long-term effects on anthropometric indicators of growth. 7-10 A familial patterning to A. lumbricoides infection has often been noted, 11-13 but there have been no detailed genetic epidemiologic analyses of susceptibility to this important helminthic infection.

While we are aware of no published study of the genetic determinants of susceptibility to A. lumbricoides in humans, a number of genetic studies of other parasitic helminths have been conducted using animal models such as inbred strains of mice. 14-16 Much of this work has been devoted to simply determining the presence or absence of a genetic component to the trait, without further refinement of the genetic architecture of the traits. For example, several studies have compared inbred strains of mice to infer genetic influences on a variety of immune response parameters. 14 Inbred strains of mice show significant differences in resistance to infection with, and abilities to mount an antibody response to, infection with the cestode Taenia taeniaeformis. 17,18 Different strains of inbred mice have also been shown to have distinct eosinophil responses to infection with Trichinella spiralis 19 and Necator americanus, 20 and to clear resulting worm burdens at different rates. 21

Studies of pedigrees of larger mammals have provided further evidence for significant genetic influences on helminthic infections. Dargie 22 has shown that sheep may be selected for resistance to gastrointestinal nematodes. Quantitative genetic studies in cattle have shown that helminthic worm counts and egg counts are significantly heritable, withheritabilities ranging as high as 0.93. 23,24

Studies of the genetic influences on helminthic infection in humans are limited. A familial or household patterning to helminth loads has been noted frequently, 13,25-27 but specific investigations of the influence of genetic factors on such patterning have been limited. The major deficiency of such epidemiologic examinations for subsequent genetic analysis has been the use of sampling designs that are inadequate for separating out genetic and shared environmental sources of aggregation. The optimal design for genetic analyses is to have large extended pedigrees crossing multiple households. This design allows for discrimination between the effects of shared environment (as assessed by common household) and genetics, but has rarely been used in genetic studies of susceptibility to parasitic infections.

Several association studies have suggested that genetic factors may influence susceptibility to helminthic infections. An analysis of A. lumbricoides worm loads in Nigerian children suggested the involvement of the major histocompatibility complex (MHC) in determining resistance to infection. 28 Similarly, the MHC has been associated with severity of outcome in schistosomiasis. 29 Although compelling, such association studies of candidate genes have potentially major statistical problems that can lead to false-positive results. Family-based linkage and association studies will be far more reliable than population-based association analyses for detecting the effects of specific loci on resistance/susceptibility to infection.

In one of the few pedigree-based studies of the potential role of genetics in helminthic infection, Smith and others 30 evaluated a number of transmission models for the diagnostic measure of egg counts in Strongyloides fuelleborni infection in Papua New Guinea. These investigators were unable to identify any genetic components to egg count variation, but the pedigrees were far from optimal for analysis. Only 177 individuals distributed among 47 pedigrees were used, indicating that both pedigree size and depth were extremely limited.

Quantitative genetic analysis of data on hookworm burden in a population from rural Zimbabwe demonstrated the presence of significant genetic effects on susceptibility to infection. 31 Quantitative measures of hookworm eggs per gram of feces, as determined by the Kato-Katz thick smear technique, were available for 279 individuals distributed across 62 pedigrees, and 10 independent individuals. A variance
decomposition approach showed that the heritability (± SE) of hookworm load was 0.37 ± 0.06 (P < 0.0001) in this population, indicating that 37% of the variation in hookworm eggs per gram of feces was attributable to genetic factors.\(^{31}\)

Segregation analysis techniques have been applied successfully in a study of human susceptibility to *Schistosoma mansoni*.\(^{32}\) Despite a limited sample (269 individuals in 20 pedigrees), Abel and others\(^{33}\) were able to detect a codominant major gene accounting for a substantial proportion of the variation in infection intensity. The investigators suggested that this gene also influenced post-treatment reinfection patterns, but this was not addressed explicitly. The gene detected by segregation analysis has recently been localized to chromosome five.\(^{34}\)

There have been no detailed genetic studies of susceptibility to infection with *A. lumbricoides* in human populations. The purpose of this study was to assess the genetic determinants of susceptibility to infection with *A. lumbricoides* in an endemic population from eastern Nepal.

**Population, Materials, and Methods**

**Study population.** The focal population for this study was the Jirel population of eastern Nepal. The Jirels reside in the Jiri region of Dolakha district approximately 190 km east of the capital city of Kathmandu. The Jirels are a Tibeto-Burman language speaking group, and ethnohistorical accounts and population genetic studies suggest that the Jirels are a hybrid population derived from Sunwars and Sherpas\(^{34}\) (Blangero J. 1987. *Selective Neutrality of Dermatoglyphic Variation*. Doctoral dissertation. Department of Anthropology, Case Western Reserve University, Cleveland, Ohio). The population has been the subject of extensive population genetic and genetic epidemiologic study for the last 14 years and has this has resulted in detailed knowledge of the pedigree structure of the group.\(^{34,45}\)

**Family information.** Household surveys were performed to obtain demographic, pedigree, and socioeconomic data from each household. The interview began with an explanation of the study and of the samples to be taken from participants. Consent forms written in Nepali were presented to each household member to be read by the individual, or to be read to the individual by one of the research assistants. Questions were answered at this time, and informed consent was obtained from all individuals who participated in the study. The protocol and consent form were approved by the Internal Review Board of the University of Texas Health Science Center at San Antonio and the Nepal Health Research Council in Kathmandu.

For each household member, sex and birth date were recorded. Additionally, birthplace, clan, and parental names were noted for all individuals. Clans are patrilineally inherited and provide a supplementary means of subgroup identification. Demographic and pedigree data are maintained in PEDSYS, a pedigree-based data management system.\(^{46}\)

**Worm burden phenotypes.** We determined both egg counts and total worm burden in a 96-hr collection period. Every study participant was given the recommended dosage (400 mg) of albendazole (Zentel\(^{47}\); Smith Kline Beecham, London, United Kingdom) to effect elimination of worms. Albendazole requires only a single treatment for effective treatment of *Ascaris* infection.\(^{37-49}\) Individuals younger than three years of age, pregnant females, and individuals with acute illness were excluded from treatment. Egg counts were determined by the Kato-Katz thick smear method from a pretreatment fecal sample. Two egg counts were performed for each specimen with the results averaged to generate the egg count phenotype for each individual.

All study participants were instructed to save stools for 96 hr following the administration of the anthelminthic drug. A prelabeled, two-liter, plastic container with lid a was given to each subject. Defecation was performed directly into the receptacle and samples were retrieved by research assistants on a daily basis for determination of worm counts.

*Ascaris* worm counts were obtained as follows. After addition of water to feces in a shallow dish, *Ascaris* worms were removed with forceps, tipped onto a mesh screen, washed in tap water, and separated by sex before being counted and weighed. Worm burden (total number of worms and sex-specific counts) and worm biomass (total worm weight) were recorded.

**Sample.** The longitudinal study enrolled subjects over a two-year period. In the first year, 1,261 individuals were sampled, including 659 females and 602 males. The mean age at examination was 25.4 years (range = 3–85 years). Of the 1,261 individuals, 1,261 had egg count data and 1,007 had complete 96-hr stool samples. To assess reinfection patterns and to equalize the exposure period, 1,002 individuals were re-examined for egg counts in the second year with 965 providing complete 96-hr stool samples. A total of 910 individuals had complete 96-hr stool samples available for both the first and second years.

**Pedigree structure of the sample.** Pedigree links were established between all individuals in the data set based on the detailed family information collected during the course of the current study and the 14 previous years of research with the population. The extensive in-depth pedigree information allowed reconstruction of a single pedigree containing all 1,261 individuals. Of these individuals, 157 are founders (i.e., individuals whose parents are unknown or are not needed to determine additional pedigree links), and the remaining 1004 individuals are members of 521 sibships. The mean sibship size is 2.12 individuals, with a range of 1–7 individuals sampled. Table 1 shows the numbers and types of relative pairs present in the Jirel sample.

**Analytic methods.** To assess the potential for susceptibility to be due to host-specific factors, we calculated the correlations for the worm burden phenotypes between the two time points. For the two quantitative measures, egg counts and worm counts, we calculated correlation coefficients simultaneously correcting for age and sex effects. We used a variance decomposition approach to determine the proportion of variation present in the ascaris phenotypic traits attributable to genetic factors. This general approach allows partitioning of the variance in a trait among components attributable to genetics, to systematic environmental components, and random environmental components, while simultaneously incorporating covariate effect. All analyses were performed using the computer package SOLAR,\(^{50,51}\) which uses the program FISHER\(^{52}\) as the variance component engine. Because of the availability of both pedigree and
TABLE 1

Relative pairs in the Jirel sample and degree of relationship

<table>
<thead>
<tr>
<th>Relationship</th>
<th>n</th>
<th>Degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent-offspring</td>
<td>1,365</td>
<td>1</td>
</tr>
<tr>
<td>Sibling</td>
<td>1,075</td>
<td>1</td>
</tr>
<tr>
<td>Grandparent-grandchild</td>
<td>580</td>
<td>2</td>
</tr>
<tr>
<td>Avuncular</td>
<td>1,608</td>
<td>2</td>
</tr>
<tr>
<td>Half-sibling</td>
<td>164</td>
<td>2</td>
</tr>
<tr>
<td>Double first cousin</td>
<td>34</td>
<td>2</td>
</tr>
<tr>
<td>First cousin</td>
<td>2,264</td>
<td>3</td>
</tr>
<tr>
<td>Grand avuncular</td>
<td>927</td>
<td>3</td>
</tr>
<tr>
<td>Half avuncular</td>
<td>458</td>
<td>3</td>
</tr>
<tr>
<td>Great grandparent-grandchild</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Fourth degree relative</td>
<td>4,702</td>
<td>4</td>
</tr>
<tr>
<td>Fifth degree relative</td>
<td>5,715</td>
<td>5</td>
</tr>
<tr>
<td>Sixth degree relative</td>
<td>3,812</td>
<td>6</td>
</tr>
<tr>
<td>Seventh degree relative</td>
<td>2,393</td>
<td>7</td>
</tr>
<tr>
<td>Eighth degree relative</td>
<td>903</td>
<td>8</td>
</tr>
<tr>
<td>Ninth degree relative</td>
<td>26</td>
<td>9</td>
</tr>
<tr>
<td>Other higher degree relative</td>
<td>181</td>
<td>—</td>
</tr>
<tr>
<td>Total relative pairs</td>
<td>26,213</td>
<td>—</td>
</tr>
</tbody>
</table>

For each Ascaris burden phenotype, we estimated both the heritability \( h^2 \), the relative proportion of the total phenotypic variance of worm burden that is attributable to genetic factors), and the proportion of variance attributable to the effects of shared environment as assessed here by common household \( c^2 \). Separate analyses were performed for each phenotype for each year and for the sum (or average) of the two years.

We considered a series of nested models and determined the best fitting model given the observed data using likelihood ratio tests. The most general model allowed for both genetic and shared environment (i.e., common household) effects. The shared environmental model did not allow for variation attributable to a genetic component while the genetic model excluded effects of shared environment. Finally, a sporadic model that lacked both genetic and shared environmental effects was fit to the data. Hypothesis tests were performed using likelihood ratio statistics that compared the competing models. In all cases we simultaneously estimated covariate effects including sex and sex-specific age and age\(^2\) effects. Maximum likelihood estimates of parameters were obtained by numerical maximization of the pedigree likelihood.

RESULTS

The prevalence of A. lumbricoides infection in the Jirel population is moderately high. During the first year of the study, 27.2% of the 1,261 individuals sampled for egg counts were positive for roundworm infection. One year after treatment with albendazole, the prevalence of roundworm infection among the 1,002 resampled individuals was remarkably similar at 24.2%. Figure 1 shows the age-specific prevalences of Ascaris infection in the Jirel population by year. In both years, there are no significant differences among age classes.

Figure 2 shows age-specific intensity of infection as indicated by geometric mean egg counts. The age profiles for each year are parallel, although egg counts were significantly \((P < 0.01)\) reduced in the second year of study for all but the youngest age group. The youngest individuals showed the highest intensities of infection. For the total population, the geometric mean egg count for the first year (1,452.9 eggs per gram of feces [egp]) was significantly \((P < 0.01)\) higher than that observed in the second year (882.2 epg).

Worm counts were determined from the 96-hr stool samples. Individuals who were infected with Ascaris expelled an average of 2.37 worms (SD = 4.33); the largest number of worms expelled by an individual was 43. The pattern of expulsion was similar in the second year with an average of 2.67 worms expelled (SD = 4.48), and a maximum expulsion of 37 worms. Figure 3 shows the distribution of worm counts in infected individuals obtained from the 96-hr collection of stools following treatment with albendazole. Both years are combined in the figure.
Phenotypic correlations. The phenotypic correlations between the two time periods were positive for all worm burden phenotypes, indicating that certain individuals appear to be predisposed to having higher susceptibility to *Ascaris* infection. The Pearson product correlation between the two time points was $0.453$ ($P < 0.0001$) for egg counts, and $0.787$ ($P < 0.0001$) for worm counts. These high intra-individual correlations provide the initial evidence that there is great potential for either genetic factors or consistent environmental factors to be influencing susceptibility to *Ascaris* infection.

The correlations between the two worm burden measures were also significant within years. The correlation between egg and worm count was $0.685$ ($P < 0.0001$) in the first year and $0.625$ ($P < 0.0001$) in the second year.

Genetic analyses. The genetic analyses are designed to explicitly differentiate between the two hypotheses resulting from the assessment of phenotypic correlations. They allow evaluation of the relative effects of genetic factors and environmental factors on the traits related to *Ascaris* infection. Three measures of worm burden were analyzed, including eggs per gram of feces, direct worm counts, and worm biomass (weight). For all the traits, there is unequivocal evidence for a strong genetic component accounting for between 30% and 50% of the variation in worm burden. The results for each of the assessed phenotypes are shown in Table 2.

**Eggs per gram of feces.** The first trait subjected to genetic analysis was eggs per gram of feces as determined in the first year of the study. This phenotype represents the pretreatment state since the egg counts were determined prior to administration of albendazole. The heritability of eggs per gram of feces was $0.291$ ($P < 0.0001$) in the first year, indicating that genetic factors accounted for close to 30% of the variation in worm loads as assessed by initial egg counts. Common household effects accounted for an additional 6% ($P = 0.0016$) of the variation. The first year data represent worm loads after variable exposure times, and thus may have hidden variation which could diminish the genetic signal.

The data determined in the second year represent infection after one year of exposure subsequent to anthelmintic treatment. In the second year, the heritability of eggs per gram of feces was increased to $0.399$ ($P < 0.0001$), indicating that approximately 40% of the variation in worm loads one year after anthelmintic treatment could be attributed to genetic factors. Common household effects were determined to be responsible for an additional 11.3% ($P < 0.0001$) of the variation in worm loads. When egg counts from the first and second years were averaged, the heritability of the resultant long-term worm load measure was $0.390$ ($P < 0.0001$) and common household effects accounted for 13.2% of the observed variation ($c^2 = 0.132, P < 0.0001$).

**Worm counts.** Total worm counts determined after anthelmintic treatment showed similar heritabilities to those observed for the egg count data. In the first year of the study, genetic factors accounted for 36% of the variation in worm count ($h^2 = 0.359, P < 0.0001$), while shared environmental (i.e., common household) factors did not account for a significant proportion of the variation ($c^2 = 0.030, P = 0.0923$). Reflecting the same pattern observed for the egg count data, the worm counts from the second year exhibited a higher heritability with the point estimate occurring at $0.412$ ($P < 0.0001$); significant environmental effects were also observed ($c^2 = 0.223, P < 0.0001$). A composite measure of long-term worm burden was generated by summing individual worm burdens across the two years. This composite measure was significantly heritable ($h^2 = 0.441, P < 0.0001$), and also exhibited significant shared environmental effects ($c^2 = 0.146, P < 0.0001$).

**Worm weight.** Our final measure of worm burden was total worm weight of all worms collected from an individual subsequent to treatment. This measure of worm burden was also heritable in the Jirel population, with approximately 34% of the variation in the first year measures being attributable to genetic factors ($h^2 = 0.335, P < 0.0001$), and approximately 48% of the variation in the second year measures being attributable to genetic factors ($h^2 = 0.477, P < 0.0001$). There was no evidence for significant common household effects on the first year worm weight measures ($c^2 = 0.034, P = 0.0584$), but there was a significant common household effect on the second year worm weights ($c^2 = 0.125, P < 0.0001$). When the worm weights were summed over the two years, we again obtained a significant heritability ($h^2 = 0.368, P < 0.0001$), indicating that approximately 37% of the variation in total worm weight was attributable to genetic factors. Common household effects were estimated at 11.2% for total worm weight summed over the two years ($c^2 = 0.112, P < 0.0001$).

### Table 2

<table>
<thead>
<tr>
<th>Trait</th>
<th>$h^2$</th>
<th>$P$</th>
<th>$c^2$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eggs per gram (Year 1)</td>
<td>0.291</td>
<td>$&lt;0.0001$</td>
<td>0.061</td>
<td>0.0016</td>
</tr>
<tr>
<td>Eggs per gram (Year 2)</td>
<td>0.399</td>
<td>$&lt;0.0001$</td>
<td>0.113</td>
<td>$&lt;0.0001$</td>
</tr>
<tr>
<td>Worm burden (Year 1)</td>
<td>0.359</td>
<td>$&lt;0.0001$</td>
<td>0.030</td>
<td>0.0923</td>
</tr>
<tr>
<td>Worm burden (Year 2)</td>
<td>0.412</td>
<td>$&lt;0.0001$</td>
<td>0.223</td>
<td>$&lt;0.0001$</td>
</tr>
<tr>
<td>Worm burden (Sum)</td>
<td>0.441</td>
<td>$&lt;0.0001$</td>
<td>0.146</td>
<td>$&lt;0.0001$</td>
</tr>
<tr>
<td>Weight (Year 1)</td>
<td>0.335</td>
<td>$&lt;0.0001$</td>
<td>0.034</td>
<td>0.0584</td>
</tr>
<tr>
<td>Weight (Year 2)</td>
<td>0.477</td>
<td>$&lt;0.0001$</td>
<td>0.125</td>
<td>$&lt;0.0001$</td>
</tr>
<tr>
<td>Weight (Sum)</td>
<td>0.368</td>
<td>$&lt;0.0001$</td>
<td>0.112</td>
<td>$&lt;0.0001$</td>
</tr>
</tbody>
</table>
DISCUSSION

Relatively few genetic studies of susceptibility to infectious diseases have been conducted in human populations. However, recent developments in statistical and molecular genetics have created a research environment where it is now possible to explore in detail the genetic and environmental factors influencing susceptibility to a broad range of infectious diseases. The future discovery of the genes involved in infectious disease processes has exciting implications for the development of novel prevention and intervention mechanisms.\(^{53}\) Many of the parasitic helminthic infections are well suited to genetic analysis because of ease of assessment and quantification. Since they are important diseases in and of themselves, results of genetic studies of helminthic infections may pave the way for significant improvements in the control of these diseases.

Infection with *A. lumbricoides* continues to be a significant health problem throughout the world. The prevalence data presented in this paper demonstrate that roundworm infection occurs at a high level in the Jirel population. Soil contamination is evidently high, since the prevalence found one year after treatment is nearly identical to that found in the initial assessment. The high within-individual correlations indicate that there is individual predisposition to infection with *A. lumbricoides*.

The hypothesis that this individual predisposition is primarily attributable to genetic factors is strongly supported by the heritability estimates for all three measures of worm burden at both points in time. For all traits, there is unequivocal evidence for a genetic component accounting for between 30% and 50% of the variation in worm burden. The results are remarkably consistent across phenotypes and across years of the study.

There is also substantial evidence for shared environmental (i.e., common household) factors influencing worm burden. These shared environmental effects account for between 3% and 13% of the total phenotypic variance. This relatively small effect of shared environment can also be seen in the correlations between spouses (who are unrelated but live in the same household). For all the roundworm burden traits, this correlation is low (\(r_w \approx 0.08\)).

The heritability estimates are consistently higher for the second year data compared with those evaluated for the first year data. This is also true of the common household effects variable. The improved resolution probably reflects a decrease in error variability that may be attributable to variable length of exposure to risk of infection in the first year data. The data from the second year reflect endpoints uniformly assessed one year after anthelmintic treatment.

The results presented here document for the first time the remarkable consistency of the re-
sults through time, and across different measures of the burden phenotype strongly indicate their validity, and suggest the utility of future evaluations focused on identifying the specific genes responsible for these sizable genetic effects.

Acknowledgments: The generous cooperation of the Jirel people with the Jiri Helminth Project is gratefully acknowledged. We thank Suman Jirel, Chhatra Jirel, Dev Bahadur Jirel, Chandra Jirel, Mithe Jirel, Kamala Khadka, Ram Upreni, Robin Singh Shrestha, and Cheryl Raindl for dedicated work on the project, and Michael Gottlieb for encouragement and support.

Financial support: This research was supported by National Institutes of Health grant R01 AI-37091, the Southwest Foundation for Biomedical Research Founder’s Council, and the Southwest Foundation for Biomedical Research. The statistical methods were developed under NIH Grant MH59490.

Authors’ addresses: Sarah Williams-Blangero, Edward S. Robinson, and John Blangero, Department of Genetics, Southwest Foundation for Biomedical Research, PO Box 761549, San Antonio, TX 78245-0549. Janardan Subedi, Department of Sociology, Anthropology, and Gerontology, Miami University, Oxford, OH 45056. Ram Prasad Upadhayay, Deepak Manral, Dev Raj Rai, and Bharat Jha, Institute of Medicine, Tribhuvan University, Kathmandu, Nepal and the Jiri Helminth Project, Jiri, Nepal.

REFERENCES

of mice to various levels of infection with the eggs of *Taenia taeeniiformis*. *Int J Parasitol* 14: 541–546.


