Two Quantitative Trait Loci Influence Whipworm (*Trichuris trichiura*) Infection in a Nepalese Population

Sarah Williams-Blangero,¹ John L. VandeBerg,¹,² Janardan Subedi,³ Bharat Jha,⁴ Tom D. Dyer,¹ and John Blangero¹

¹Department of Genetics, Southwest Foundation for Biomedical Research, and ²Southwest National Primate Research Center, San Antonio, Texas; ³Department of Sociology and Gerontology, Miami University, Oxford, Ohio; ⁴Tribhuvan University Institute of Medicine, Kathmandu, Nepal

**Background.** Whipworm (*Trichuris trichiura*) infection is a soil-transmitted helminth infection that affects >1 billion people. It is a serious public health problem in many developing countries and can result in deficits in growth and cognitive development. In a follow-up study of significant heritability for whipworm infection, we conducted the first genome scan for quantitative trait loci (QTL) influencing the heritability of susceptibility to this important parasitic disease.

**Methods.** Whipworm egg counts were determined for 1253 members of the Jirel population of eastern Nepal. All individuals in the study sample belonged to a single pedigree including >26,000 pairs of relatives that are informative for genetic analysis.

**Results.** Linkage analysis of genome scan data generated for the pedigree provided unambiguous evidence for 2 QTL influencing susceptibility to whipworm infection, one located on chromosome 9 (logarithm of the odds ratio [LOD] score, 3.35; genomewide \( P = .0138 \)) and the other located on chromosome 18 (LOD score, 3.29; genomewide \( P = .0159 \)). There was also suggestive evidence that 2 loci located on chromosomes 12 and 13 influenced whipworm infection.

**Conclusion.** The results of this first genome scan for *T. trichiura* egg counts provides new information on the determinants of genetic predisposition to whipworm infection.

It is estimated that whipworm (*Trichuris trichiura*) infection affects >1 billion people [1]. High whipworm loads are associated with severe consequences, such as rectal prolapse and dysentery, that have an immediate and obvious influence on the health of the infected individual [2, 3]. However, even infections that are classified by the World Health Organization (WHO) as moderate or nonsevere can have significant long-term impacts on health [3, 4].

The association between whipworm infection and malnutrition has been documented in numerous human populations throughout the developing world [3, 5–9]. Although anthelmintic treatments may help achieve some recovery from the consequences of infection-associated malnutrition [6, 10, 11], the deficits in childhood growth and development associated with whipworm infections have long-term implications [8, 12]. Stunting is evident even in children with low worm loads [13].

*Trichuris* infection has been associated with iron deficiency anemia, especially when other intestinal worm infections are present, in a number of studies [3, 14–16]. Vitamin A deficiency has also been associated with whipworm infection [17].

One of the most insidious effects of whipworm infection is its effect on cognitive development [2, 3]. A study of Filipino children who were 7–18 years of age found that whipworm infection was associated with deficiencies in verbal fluency [15]. Similarly, a study conducted in Jamaica found that whipworm infection had significant effects on cognitive functioning in children 9–12
years of age [18]. The effects of deficits in cognitive development on the ability to learn may have lifetime consequences [3, 4].

Like other soil-transmitted helminthic infections, whipworm infection is characteristically overdispersed in human populations, with a small proportion of available hosts harboring the majority of the parasitic worms [19–21]. A genetic basis to this apparent predisposition to whipworm infection in a small percentage of the population has been supported by a number of studies [22]. Our previously published study demonstrated significant heritabilities for susceptibility to infection with T. trichiura in 2 independent Asian populations [23].

Susceptibility to whipworm infection was significantly heritable in a Han Chinese population from Jishan Island in the Jiangxi Province of the People’s Republic of China. Approximately 86% of the population was infected with T. trichiura. A quantitative genetic analysis of the data from the Jishan Island population revealed that 29% of the variation in egg counts was attributable to genetic factors [23].

The heritability of susceptibility to whipworm infection was also assessed in the Jirel population of eastern Nepal, the focal population for the analyses reported in the present article. The analyses of data from >1200 individuals resulted in a significant heritability of whipworm eggs per gram of feces, which indicated that 28%–36% of the variation was attributable to genetic factors [23].

Our previous analysis of susceptibility to whipworm infection [23] indicated that genetic factors are involved in determining differential susceptibility to T. trichiura. However, the quantitative genetic approaches that were used cannot identify the specific genetic factors involved in determining the observed overdispersion of whipworm infections. The purpose of this article is to present the results of the first genomewide scan for quantitative trait loci (QTL) that influence egg counts associated with T. trichiura infection. The data for this analysis were generated by the Jiri Helminth Project, a long-term genetic study of soil-transmitted helminth infections in the Jirel population of eastern Nepal.

**MATERIALS AND METHODS**

### Study Population

The focal population for the present study was the Jirel ethnic group of eastern Nepal. This Tibeto-Burman language–speaking group is a hybrid population descended from Sherpa and Sunwar ancestors. Members of the population reside in 9 villages in the Jiri region of Dolakha district. The region is situated at an average elevation of 7500 ft and is located ~190 km east of the capital city of Kathmandu. The study sample reported here comprised individuals from 6 of the 9 Jirel villages.

Individuals were recruited into the study according to a protocol that was approved by the institutional review board at the University of Texas Health Science Center at San Antonio and the Nepal Health Research Council. Data on whipworm infection were available for a total of 1253 individuals, including 601 males and 652 females. Members of the study sample were 3–85 years of age (average age, 25.4 years).

The Jirel population has been the subject of population genetic and genetic epidemiologic research since 1985. As a result, there is a vast amount of available information on the genetic characteristics of the population and the kinship relationships among members of the population. Pedigree information was collected according to the protocol described elsewhere by Williams-Blangero and Blangero [24]. The detailed information available on the relationships among the individuals participating in the study has allowed the assignment of all individuals in the sample to a single, highly complex, extended pedigree, which is extremely powerful for genetic analysis [24]. The pedigree used for the analyses described in the present article includes >26,000 pairs of relatives that are informative for genetic analysis [23].

### Sampling for Quantitation of Egg Counts

Trichuris egg counts were defined as the number of eggs per gram of feces from small fecal samples (sample weight, ~5 g). Of the 1253 study participants, 1186 (94.7%) provided fecal samples on 2 consecutive days. All samples were assessed within 1 h of their delivery to the field laboratory by a medical technologist. Egg counts were determined using the Kato-Katz thick-smear technique with the use of a kit developed in Nepal, according to the methods outlined by WHO and implemented according to the protocol defined by WHO [25]. Two egg counts were performed for each sample, and an average count was determined from the results. There was no statistically significant difference in egg counts between days. For individuals for whom only one sample was available, the average value of the 2 available egg counts was used as the measurement. The quantitative egg count was the trait used for analysis.

### Analytic Genotyping Methods

Genotyping was performed using the ABI Prism Linkage Mapping Set (version 2; Applied Biosystems). All marker typing was done using automated sequencers (ABI DNA Sequencer Models 377 and 3100; Applied Biosystems). A total of 371 dinucleotide short tandem repeat markers evenly distributed across the 22 autosomes were characterized for each individual in the sample, resulting in a 10-cM genome scan for linkage mapping.

### Statistical Genetic Methods

#### Error checking.

SimWalk2 [26] was used to check for the Mendelian consistency of the genotypic data. This program uses Markov chain Monte Carlo and simulated annealing algorithms to determine the probabilities of mistyping for each genotype.

#### Allele frequency estimation.

Allele frequencies were estimated using maximum likelihood techniques as implemented in
the computer program SOLAR [27]. Incorporating the pedigree structure in the estimation of allele frequencies is critical, because estimations of gene frequency can be significantly biased if the relationships among individuals in the sample are not taken into account [28].

We use a variance components linkage method, as implemented in the computer program SOLAR [27], for all analyses of the genome scan data. We estimated all the elements of a location-specific identity-by-descent (IBD) probability matrix jointly, by use of a Markov chain Monte Carlo method incorporated into the program LOKI [29]. Multipoint IBD matrices were then imported into SOLAR for the linkage analyses.

**Linkage analysis.** The standard likelihood framework for variance components was used for all estimations of parameters and testing of hypotheses. We tested the null hypothesis that the additive genetic variance due to the $i$th QTL equals zero (denoting no linkage), by comparing the likelihood of the restricted model with that of a model in which the variance due to the $i$th QTL was estimated. The difference between the 2 log$_{10}$ likelihoods provided the logarithm of the odds ratio (LOD) score, which is a measurement of the support for the hypothesis of linkage over that of “no linkage” at a particular chromosomal location. $P$ values were determined using a test statistic that is twice the difference in the log, likelihoods of the 2 models and that is asymptotically distributed as a 12:12 mixture of a $\chi^2$ variable with a point mass at zero [30]. Because of the large number of tests required to scan the genome by use of this linkage procedure, we calculated analytical genomewide $P$ values [31]. Thus, all statistical results of linkage are provided with experiment-wide levels of significance.

Because the data on whipworm infection tended to be leptokurtic, we used robust LOD scores [32, 33] when evaluating the log-transformed *Trichuris* egg counts [ln(egg + 1)] Word Functions menu in ADEPT to insert roman function properly. (v3)/$H_9273^2$,$H_9273^2$/ to avoid the potential problem of nonnormality leading to excessive type I error [34].

Within each model, we corrected for multiple covariates as potential confounders. For the current analyses, we simultaneously controlled for the effects of sex, age, age$^2$, sex $\times$ age, sex $\times$ age$^2$, and whether or not an individual had daily access to a latrine (as a marker of potential economic-related hygiene) on an ln-transformed *Trichuris* egg count per gram of feces.

**RESULTS**

The prevalence of whipworm infection in the Jirel population was 14.4%. The heritability (or familiality) of the whipworm egg count in this sample of individuals in the Jirel population is highly significant ($h^2 = 0.38 \pm 0.06$; $P = 6.5 \times 10^{-17}$). The genomewide linkage analyses performed for the 1253 members of the Jirel population localized 2 significant QTL influencing susceptibility to whipworm eggs per gram of feces. Figure 1 presents the string plot of the linkage results for the complete genome scan. Two regions demonstrated genomewide significance, with LOD scores of 3.35 and 3.29 (for chromosomes 9 and 18, respectively). Two additional suggestive signals, with LOD
scores of 2.28 and 2.04, respectively, were found on chromosomes 12 and 13.

Figure 2 presents the linkage results for chromosome 9 in more detail. The linkage peak on chromosome 9 occurred at 2 cM. The LOD score for this QTL was 3.347 (genomewide \( P = .0138 \); nominal \( P = 4.3 \times 10^{-5} \)). The chromosomal region of the peak was 9p24. The 1-LOD support interval for the peak spanned from the P terminus to 7 cM, corresponding to an \( \sim 14 \)-Mb physical region on chromosome 9.

Figure 3 presents similar results for chromosome 18. The peak LOD score on chromosome 18 occurred at 31 cM. The approximate chromosomal location of the peak signal was 18p11. The LOD score for this QTL reached 3.290, which was significant (genomewide \( P = .0159 \); nominal \( P = 5 \times 10^{-5} \)). The 1-LOD support interval for this linkage signal was larger than the one on chromosome 9 and spans from 22 to 43 cM. This relatively large genetic region spanned a relatively small physical region of \( \sim 6 \) Mb on chromosome 18.

The linkage peak on chromosome 13 occurred at 37 cM (LOD score, 2.276). This result is suggestive of a gene influencing susceptibility to whipworm infection (genomewide \( P = .206 \); nominal \( P = .0006 \)). Thus, in a given genome scan, we would expect 0.21 occurrences of a LOD score at least this large by chance. Similarly, the other suggestive evidence of a QTL influencing susceptibility to whipworm infection is on chromosome 12 at 191 cM. This second linkage peak had a LOD score of 2.042 (genomewide \( P = .373 \); nominal \( P = .0011 \)).

Interestingly, there was no overlap between the QTL influencing whipworm egg counts and the genes localized for infection with *Ascaris lumbricoides* [35]. This finding suggests that the major loci influencing helminthic burden are probably infection specific.

**DISCUSSION**

The present study reports the first genomewide linkage scan for QTL influencing susceptibility to infection with *T. trichiura*. Our findings of \( \geq 2 \) significant QTL regions is consistent with the previously reported significant heritability of susceptibility to whipworm infection in this population [23]. Localization of the QTL involved in the determination of a complex disease trait is the first step toward gene identification. Given the objective prioritization that such linkage analyses proffer, we now plan to dissect these genomic regions by use of a combined approach involving in silico data mining and linkage disequilibrium mapping, to nominate positional candidate genes for exhaustive resequencing and formal detection of the most likely functional variants [36]. Once the specific individual genes influencing whipworm infection have been identified, the information can be used to potentially identify novel drug and vaccine targets [37, 38]. The analyses reported here have unambiguously localized 2 QTL influencing whipworm infection, and they have also provided suggestive evidence for 2 additional QTL that may be involved in determining differential susceptibility to whipworm infection.

Preliminary investigation into the potential positional candidate genes within the 1-LOD support interval of these 2 QTL identified several interesting possibilities. Within the QTL
region on chromosome 9, there are ~161 known or predicted genes, most of which have unknown functions. One potential candidate is the JAK2 gene, which is found near the linkage peak (at ~4.9 Mb). Activation of the Janus-activated kinase 2 (JAK2)/STAT1 signaling pathway is suppressed in Leishmania-infected macrophages [39], and JAKE2 appears to be involved in the nitric oxide killing response to parasitic infection [40]. Also in this immediate region on chromosome 9 can be found the PDCD1LG1 (also known as B7H1) genes, which are known to be involved in the stimulation of T cell response and which have a particularly strong effect on the production of interleukin (IL)–10 [41]. IL-10 is well known to be critically involved in playing a role in Trichuris infection in mice [42]. In addition, in humans, levels of IL-10, among other cytokines, have been shown to be factors predictive of the susceptibility to whipworm infection [43]. The whipworm-susceptibility QTL on chromosome 18 spans ~6 Mb and contains 86 known and predicted genes. One obvious candidate gene in this region is RALBP1 (or RIP1), which is involved in the direction of cell death after stimulation by pathogens [44]. Overall, as with most genomic regions, the 1-LOD support intervals contain a number of genes with some known function, but little is known about most of the genes. Thus, it will be necessary to perform additional studies to probe these genes for functional signals related to whipworm susceptibility, either by looking for associations with sequence variants or by combining such studies with expression-based studies related to worm infection.

Whipworm infection has a huge impact on public health, affecting >1 billion people and resulting in the loss of millions of years of disability-adjusted life-years annually [3, 45]. Numerous studies have shown that individuals are readily reinfected after receiving anthelmintic treatment, indicating that there is little acquired host immunity to whipworm [19, 46, 47]. As clearly noted by Bethony et al. [2], whipworm infections are diseases associated with poverty and will continue to be major public health problems in the developing world until substantial improvements in standard of living are achieved in these areas. The long-term efficacy of mass deworming programs will be limited because of logistical and compliance problems, as well as the emergence of resistance in the parasites [2, 3]. There is a critical need for an improved understanding of the biological determinants of differential susceptibility to helminthic infections. The results of the present study provide the first step toward the identification of specific genes influencing differential susceptibility to whipworm infection.

Acknowledgments

The generous cooperation of the Jirel people is gratefully acknowledged. We thank the staff of the Jiri Helminth Project for their contributions to this research. We also thank Angie Olson, Liz Rainwater, Mary Jo Aivaliotis, Mari Hui, Ram Upadhyay, and Cheryl Raindl for their expert technical assistance.

References