Am. J. Trop. Med. Hyg., 85(5), 2011, pp. 893–896
doi:10.4269/ajtmh.2011.11-0159

Short Report: Evidence for Genetic Susceptibility to Developing Early Childhood Diarrhea among Shantytown Children Living in Northeastern Brazil

Relana C. Pinkerton, Reinaldo B. Oriá, Jack W. Kent Jr., Anita Kohli, Claudia Abreu, Oluma Bushen, Aldo A. M. Lima, John Blangero, Sarah Williams-Blangero, and Richard L. Guerrant*

University of Virginia Center for Global Health, Charlottesville, Virginia; Texas Biomedical Research Institute, San Antonio, Texas; Institute of Biomedicine and Department of Morphology, School of Medicine, Federal University of Ceara, Forteleza, Ceara, Brazil

Abstract. To explore the genetic components of susceptibility to early childhood diarrhea (ECD), we used a quantitative genetic approach to estimate the heritability of ECD among children from two Brazilian favelas. Shared environment was used to model common exposure to environmental factors. Genetic relatedness was determined from pedigree information collected by screening household participants (n = 3,267) from two geographically related favelas located in Fortaleza, Brazil. There were 277 children within these pedigrees for whom diarrheal episodes in the first two years of life were recorded. Data on environmental exposure and pedigree relationship were combined to quantitatively partition phenotypic variance in ECD into environmental and genetic components by using a variance components approach as implemented in Sequential Oligogenic Linkage Analysis Routines program. Heritability accounted for 54% of variance in ECD and proximity of residence effect accounted for 21% ($P < 0.0001$). These findings suggest a substantial genetic component to ECD susceptibility and the potential importance of future genetics studies.

Early childhood diarrhea (ECD) is a leading cause of child mortality and morbidity globally and causes more than 1.34 million deaths per year.[1,2] Children who survive repeated episodes of diarrhea often have sustained growth shortfalls and impairments in physical fitness, cognition and school performance.[3–6] Early childhood diarrhea is associated with a wide range of microbes and pathogenesis.[7,8] However, the human genes responsible for variation in susceptibility to acquiring pathogens associated with diarrhea remain poorly explored.

Host genetics have been increasingly implicated in susceptibility to specific parasites. These parasites include *Ascaris lumbricoides*, *Cryptosporidium* spp., and *Entamoeba histolytica*.[9–14] However, little is known about the importance of host genetics for risk of developing ECD. Knowledge of the genetic determinants of host susceptibility to ECD may facilitate development of novel approaches for control of the disease.

This study applied a variance component approach to quantify the relative contribution of genetic and environmental components to risk for ECD. We used extended pedigree information, global positioning mapping (GPS) of household residence and ECD data generated for children living in two favelas in Fortaleza, Brazil. The study protocol and procedures for obtaining informed consent were approved by the University of Virginia Institutional Review Board for Health Sciences Research, the local Institutional Review Board of the University of Ceara in Fortaleza, and the Brazilian Ministry of Health.

Family information was gathered from two urban shantytowns (Parque Universitario and Gonçalves Dias) located within the capital city of Fortaleza, Ceara, in northeastern Brazil. Pregnant women in these communities were identified, and enrollment of newborns was initiated in 1989. Each child in the study was visited three times a week by a study nurse who recorded diarrheal illness, sociodemographics data, and breastfeeding information. Children who had diarrhea were visited daily until 48 hours after resolution of illness. Mothers were asked to provide detailed clinical information about the illness including stool characteristics and other symptoms such as fever, vomiting, and dehydration.

Extended pedigrees were reconstructed on the basis of information gathered for all household participants in the two favelas described according to the approach of Williams-Blangero and Blangero.[15]

The World Health Organization criteria to determine diarrhea episodes were used for our analyses. Diarrhea was defined as ≥ 3 liquid stools in the preceding 24-hour period. An episode of diarrhea was defined as lasting ≥ 1 day and separated from another episode by ≥ 2 days without diarrhea. Household locations were mapped by using GPS. Distance information was included in the variance component models to assess risk for ECD caused by geography and common living environment.

We used the Sequential Oligogenic Linkage Analysis Routines (SOLAR) program to assess the amount of variance accounted for by genetic relatedness in developing ECD. This program calculates variance in risk for ECD attributable to genetic factors. It also uses a variance component approach to partition the total phenotypic covariance ($\mathbf{\Omega}$) of the traits into additive genetic (σ^2_G) and environmental (σ^2_E) variance components according to the formula $\mathbf{\Omega} = 2\mathbf{\Phi}\sigma^2_G + \mathbf{I}\sigma^2_E$, where $\mathbf{\Omega}$ is a matrix of phenotypic covariance between pairs of persons, $2\mathbf{\Phi}$ is a matrix of expected allele sharing on the basis of degree of relationship, and $\mathbf{I}$ is an identity matrix. The environmental component includes factors that are not explicitly measured such as socioeconomic status, measurement errors, and dominance or other non-additive genetic factors. The additive heritability (h^2) of a trait represents the portion of the total phenotypic variance σ^2_P accounted for by the additive genetic variance formula $h^2 = \sigma^2_G/\sigma^2_P$.

To determine the significance of heritability, a null model in which the additive genetic variance (σ^2_G) for the trait equals zero is compared with another model, where σ^2_G is estimated by maximum likelihood methods. Because σ^2_G is tested on its boundary, the likelihood ratio test statistic is distributed as a ½:½ mixture of a chi-square distribution with one degree of freedom and a point mass at zero.

The decomposition of the phenotypic covariance is readily extended to include other random effects. We tested models

*Address correspondence to Richard L. Guerrant, University of Virginia Center for Global Health, Charlottesville, VA 22908. E-mail: rlg9a@virginia.edu

that enabled two such effects: the effect of common household (σ^2_C), and the effect of physical proximity (σ^2_{PX}) based on GPS data. The former variance component was structured by a matrix whose elements were 1 if the children shared a household or 0 otherwise. The latter variance component was structured by a proximity matrix computed as follows: for each pair of children *i,j* with ECD phenotypes, a proximity score was calculated as $px_{ij} = \exp(-d_{ij})$, where d_{ij} is the physical distance in hectometers between the children's households. Thus, the proximity score ranged from 1 for children sharing a household to effectively zero at a distance ≥ 1 km.

There were 277 children for whom ECD and GPS data were available. Children characterized for ECD belonged to 127 pedigrees, which ranged from 3 to 188 family members and had a mean (SEM) of 25.72 (2.3) members (Table 1). Pedigrees contained a mean (SEM) of 10.13 (0.87) nuclear families, which ranged from 1 to 72 nuclear families per pedigree. Each of these contained approximately equal numbers of full (mean = 2.44, SEM = 0.11), maternal (mean = 2.83, SEM = 0.11), and paternal (mean = 2.51, SEM = 0.11) sibships. Most (72.9%) pedigrees included at least one full sibship.

The degree of genetic relatedness among children in the study is delineated in Table 2. There was a high proportion (36.5%, 118 of 323 pairwise relationships) of first cousin pairs within this population. Because cousins typically lived in separate households, these and cousins of other degrees are a particularly informative group for distinguishing between genetic and environmental influences.

There was a high prevalence (89.5%, 248 of 277) of diarrhea before two years of age. More than half (50.5%, 140 of 277) the sampled children experienced ≥ 4 episodes, and 27.4% (76 of 277) had ≥ 7 or more episodes. Most mothers (82.5%, 212 of 257) had not completed elementary school. The average child was breastfed until 5.4 months of age (range = 0–24 months). Household crowding averaged 5.6 family members sleeping in a single home and ranged from 2 to 16. Average per capita income per month was approximately US $113 and ranged from none to US $755.

The eight models (A–H) used to assess the genetic versus environmental contribution to ECD are described in Table 3. A model was also run to assess the likelihood that the phenotypic variance attributed to genetics shows a more dominant or additive inheritance pattern. Variance component estimates, standard error of variance estimate, and $\log_e$ likelihood for each model are reported. The test column in Table 3 indicates which models where compared, and the *P* values refer to the probability that the models compared are not significantly different from one another; $P < 0.05$ was considered significant.

Table 1
Characteristics of pedigrees included in analysis of early childhood diarrhea, Brazil*

Characteristic	Mean	SEM	Median	Range
No. persons†	25.72	2.30	17	3–188
No. nuclear families†	10.13	0.87	7	1–72
Maternal sibship size‡	2.83	0.11	2	1–8
Paternal sibship size‡	2.51	0.11	2	1–8
Full sibship size‡	2.44	0.11	2	1–8
% full sibships‡	72.9	NA	NA	NA

* NA = not applicable.
† Pedigrees that included children with early childhood diarrhea phenotypes.
‡ Sibships of children with early childhood diarrhea phenotypes only.

Table 2
Degree of relatedness among 277 children in pedigrees with data for early childhood diarrhea, Brazil

Relationship	Coefficient of kinship	No. pairs
Sibling	1/2	80
Avuncular	1/4	13
Half-sibling	1/4	18
Half-avuncular	1/8	5
First cousins	1/8	118
First cousins once removed	1/16	27
Half first cousins	1/16	26
Second cousins	1/32	22
Second cousins once removed	1/64	6
Half second cousins	1/64	2
Double half first cousins	1/8	1
Half first cousins and second cousins	3/32	1
Half avuncular and first cousins once removed	3/16	1
Half third cousins once removed	1/512	3
Second cousins and half second cousins	3/64	2

The polygenic model (model B), which includes only additive genetic (h^2) and environmental (e^2) random effects, showed that relatedness accounts for 0.86 (SEM = 0.18) of ECD variance. This finding was highly significantly different ($P = 0.000001$) from the sporadic (random environmental) model (model A). Model C shows some evidence of a dominance effect, which interestingly absorbs all the random environmental effect (e^2). However, Model C was not significantly different from Model B ($P = 0.217637$).

Model D indicates that living in a common household has a significant effect on ECD ($P = 0.000031$), but once polygenic factors were included in Model E, the effect of household became null. The *P* value for Model E ($P = 0.500000$) indicates a non-significant difference between the combined effect of Household/Polygenic compared with Polygenic alone. Model F shows household proximity alone has a large effect size (0.74 ± 0.08), which was reduced to 0.25 ± 0.13 after polygenic factors were accounted for, as shown in Model G. The *P* value for Model G ($P = 0.022217$) indicates that proximity accounts for a significant proportion of variation in ECD, independent of polygenic factors.

The variance component analysis of ECD (Table 3) shows that a statistically significant and substantial proportion of the variation in number of episodes of diarrhea experienced in the first two years of age in these favela populations is attributable to genetic factors. Overall, genetic relatedness accounted for 0.54 ± 0.19 of the phenotypic variance of ECD, and geographic relatedness (proximity of residence) accounted for an additional 0.21 ± 0.09 ($P < 0.0001$). Furthermore, educational status of mothers, breastfeeding practices, and household crowding did not significantly account for variance in ECD once genetic relatedness and geographic proximity of residence were accounted for. Data not shown.

The study children are known to experience substantial exposure to contaminated drinking water, live in unsanitary conditions, and experience a high burden of infection.[18,16] Contaminated water sources and inadequate water storage are highly prevalent in diarrhea-endemic areas.[17] This finding, in combination with household crowding, can facilitate spread of enteric pathogens.[18,19] In our cohorts, enteroaggregative *Escherichia coli*, *Cryptosporidium*, and *Giardia* spp. are leading causes of recurrent diarrhea in children in their first two years of life and are easily spread from person-to-person.[20,21]

TABLE 3
Heritability and geographic/environmental determinants of variance in rates of early childhood diarrhea, Brazil*

Model	Variance component estimates (standard errors)					Log$_e$ likelihood	Test	P‡
	e^2	h^2	d^2	c^2	px^2			
A. Sporadic	1.00 (–)	†	†	†	†	–700.41	NA	NA
B. Polygenic	0.14 (0.18)	0.86 (0.18)	†	†	†	–689.33	B vs. A	0.000001
C. Polygenic/dominance	0.00 (–)	0.65 (0.31)	0.35 (0.31)	†	†	–689.03	C vs. B	0.217637
D. Household	0.65 (0.09)	†	†	0.35 (0.09)	†	–692.38	D vs. A	0.000031
E. Household/polygenic	0.14 (0.18)	0.86 (0.18)	†	0.00 (–)	†	–689.33	E vs. B	0.500000
F. Proximity	0.74 (0.08)	†	†	†	0.26 (0.08)	–697.01	F vs. A	0.004558
G. Proximity/polygenic	0.25 (0.13)	0.54 (0.19)	†	†	0.21 (0.09)	–687.31	G vs. B	0.022217
H. Proximity/polygenic/ dominance	0.00 (–)	0.32 (0.24)	0.47 (0.20)	†	0.21 (0.08)	–686.90	H vs. G	0.181716

*NA = not applicable.
† = Variance component constrained to 0. Variance components (standardized as fraction of phenotypic variance): e^2 = random environmental effect; h^2 = heritability; d^2 = dominance effect; c^2 = effect of shared household; px^2 = effect of proximity.
‡ P values from likelihood ratio tests.

Susceptibility to infection might be associated with disruption of intestinal function by the pathogen, leading to inflammation or tissue damage, which is also likely to be genetically influenced. Jiang and others studying single-nucleotide polymorphisms in the interleukin-8 promoter region in American travelers to Mexico found that persons with symptomatic enteroaggrgative *E. coli* had higher frequencies of AA genotype at the –251 position and increased levels of fecal interleukin-8 associated with intestinal epithelial disruption and fluid secretion.[21]

Our finding that genetic factors account for an estimated 54% of the variability in ECD suggests that further genetic studies are warranted. Several underlying phenotypes may be incorporated in the ECD phenotype subjected to analysis. For example, the variation in ECD observed in this study captures variation in acquiring infection and subsequent inflammatory or secretory responses leading to overt diarrhea. Future studies should include underlying phenotypic pathologies, such as intestinal inflammation, specific etiologies, and other phenotypic characteristics such as illness durations, related to ECD. Because co-morbidities and cumulative enteric diseases are common in children from disease-endemic areas in the developing world, a composite of genetic components of susceptibility to diverse pathogens warrant future studies of specific etiologies and pathotypes of enteric infection.

Our findings support the importance of taking the next step in investigating this topic by investing in increasingly economically feasible genome wide association or case–control studies designed to identify functional genetic variants associated with ECD in northeastern Brazil. Identification of specific genetic contributions could be used to illuminate the underlying etiologies and pathogenesis of ECD and may suggest new mechanisms to target with novel therapies.

Received March 21, 2011. Accepted for publication July 25, 2011.

Acknowledgments: We thank community field workers for collecting pedigree information and the children and family members for participating in this study.

Financial support: This study was supported in part by the National Institute of Allergy and Infectious Disease ICIDR grant U01A1026512, Support was also provided by Global Infectious Disease Research Training grant D43 TW006578.

Authors' addresses: Relana C. Pinkerton, Anita Kohli, and Oluma Bushen, University of Virginia Center for Global Health, Charlottesville, VA. Reinaldo B. Oriá and Claudia Abreu, Institute of Biomedicine and Department of Morphology, School of Medicine, Federal University of Ceara, Fortaleza, Ceara, Brazil. Jack W. Kent Jr., John Blangero, and Sarah Williams-Blangero, Texas Biomedical Research Institute, San Antonio, TX. Aldo A. M. Lima and Richard L. Guerrant, University of Virginia Center for Global Health, Charlottesville, VA, and Institute of Biomedicine and Department of Morphology, School of Medicine, Federal University of Ceara, Fortaleza, Ceara, Brazil.

REFERENCES

1. Black RE, Cousens S, Johnson HL, Lawn JE, Rudan I, Bassani DG, Jha P, Campbell H, Walker CF, Cibulskis R, Eisele T, Liu L, Mathers C; Child Health Epidemiology Reference Group Of WHO and UNICEF, 2010. Global, regional, and national causes of child mortality in 2008: a systematic analysis. *Lancet 375:* 1969–1987.
2. Kosek M, Bern C, Guerrant RL, 2003. The global burden of diarrhoeal disease, as estimated from studies published between 1992 and 2000. *Bull World Health Organ 81:* 197–204.
3. Moore SR, Lima AA, Conaway MR, Schorling JB, Soares AM, Guerrant RL, 2001. Early childhood diarrhoea and helminthiases associate with long-term linear growth faltering. *Int J Epidemiol 30:* 1457–1464.
4. Checkley W, Buckley G, Gilman RH, Assis AM, Guerrant RL, Morris SS, Molbak K, Valentiner-Branth P, Lanata CF, Black RE, 2008. Multi-country analysis of the effects of diarrhea on childhood stunting. *Int J Epidemiol 37:* 816–830.
5. Guerrant DI, Moore SR, Lima AA, Patrick PD, Schorling JB, Guerrant RL, 1999. Association of early childhood diarrhea and cryptosporidiosis with impaired physical fitness and cognitive function four-seven years later in a poor urban community in northeast Brazil. *Am J Trop Med Hyg 61:* 707–713.
6. Guerrant RL, Oria RB, Moore SR, Oria MO, Lima AA, 2008. Malnutrition as an enteric infectious disease with long-term effects on child development. *Nutr Rev 66:* 487–505.
7. Cheng AC, McDonald JR, Thielman NM, 2005. Infectious diarrhea in developed and developing countries. *J Clin Gastroenterol 39:* 757–773.
8. Steiner TS, Samie A, Guerrant RL, 2006. Infectious diarrhea: new pathogens and new challenges in developed and developing areas. *Clin Infect Dis 43:* 408–410.
9. Williams-Blangero S, Vandeberg JL, Subedi J, Jha B, Correa-Oliveira R, Blangero J, 2008. Localization of multiple quantitative trait loci influencing susceptibility to infection with *Ascaris lumbricoides*. *J Infect Dis 197:* 66–71.
10. Ramsay CE, Hayden CM, Tiller KJ, Burton PR, Goldblatt J, Lesouef PN, 1999. Polymorphisms in the beta2-adrenoreceptor gene are associated with decreased airway responsiveness. *Clin Exp Allergy 29:* 1195–1203.
11. Peisong G, Yamasaki A, Mao XQ, Enomoto T, Feng Z, Gloria-Bottini F, Bottini E, Shirakawa T, Sun D, Hopkin JM, 2004. An asthma-associated genetiv variant of STAT6 predicts low burden of *Ascaris* worm infestation. *Genes Immun 5:* 58–62.
12. Kirkpatrick BD, Haque R, Duggal P, Mondal D, Larsson C, Peterson K, Akter J, Lockhart L, Khan S, Petri WA Jr, 2008. Associations between *Cryptosporidium* infections and human leukocyte antigen class I and class II alleles. *J Infect Dis 197:* 474–478.

13. Duggal P, Haque R, Roy S, Mondal D, Sack RB, Farr BM, Beaty TH, Petri WA Jr, 2004. Influence of human leukocyte antigen class II alleles on susceptibility to *Entamoeba histolytica* infection in Bangladeshi children. *J Infect Dis 189:* 520–526.
14. Duggal P, Guo X, Haque R, Peterson KM, Ricklefs S, Mondal D, Alam F, Noor Z, Verkerke HP, Marie C, Leduc CA, Chua SC Jr., Myers MJ Jr, Leibel SR, Houpt E, Gilchrist CA, Sher A, Porcella SF, Petri WA Jr. 2011. A mutation in the leptin receptor is associated with *Entamoeba histolytica* infection in chioldren. *J Clin Invest 121:* 1191–1198.
15. Williams-Blangero S, Blangero J, 2006. Collection of pedigree data for genetic analysis in isolate populations. *Hum Biol 78:* 89–101.
16. Dillingham R, Guerrant RL, 2004. Childhood stunting: measuring and stemming the staggering costs of inadequate water and sanitation. *Lancet 363:* 94–95.
17. Copeland CC, Beers BB, Thompson MR, Fitzgerald RP, Barrett LJ, Sevilleja JE, Alencar S, Lima AA, Guerrant RL, 2009. Faecal contamination of drinking water in a Brazilian shanty town: importance of household storage and new human faecal marker testing. *J Water Health 7:* 324–331.
18. Lima AA, Moore SR, Barboza MS Jr, Soares AM, Schleupner MA, Newman RD, Sears CL, Nataro JP, Fedorko DP, Wuhib T, Schorling JB, Guerrant RL, 2000. Persistent diarrhea signals a critical period of increased diarrhea burdens and nutritional shortfalls: a prospective cohort study among children in northeastern Brazil. *J Infect Dis 181:* 1643–1651.
19. Lima AA, Guerrant RL, 1992. Persistent diarrhea in children: epidemiology, risk factors, pathophysiology, nutritional impact, and management. *Epidemiol Rev 14:* 222–242.
20. Steiner TS, Lima AA, Nataro JP, Guerrant RL, 1998. Enteroaggregative *Escherichia coli* produce intestinal inflammation and growth impairment and cause interleukin-8 release from intestinal epithelial cells. *J Infect Dis 177:* 88–96.
21. Jiang ZD, Okhuysen PC, Guo DC, He R, King TM, DuPont HL, Milewicz DM, 2003. Genetic susceptibility to enteroaggregative *Escherichia coli* diarrhea: polymorphism in the interleukin-8 promotor region. *J Infect Dis 188:* 506–511.