Gut Microbiome Changes with Acute Diarrheal Disease in Urban Versus Rural Settings in Northern Ecuador

Maria J Soto-Giron, Georgia Institute of Technology
Angela Pena-Gonzalez, Georgia Institute of Technology
Janet K Hatt, Georgia Institute of Technology
Lorena Montero, Universidad San Francisco de Quito
Maritza Páez, Universidad San Francisco de Quito
Estefania Ortega, Universidad San Francisco de Quito
Shanon Smith, Emory University
William Cevallos, Universidad Central del Ecuador
Gabriel Trueba, Universidad Central del Ecuador
Konstantinos T Konstantinidis, Georgia Institute of Technology

Only first 10 authors above; see publication for full author list.

Journal Title: AMERICAN JOURNAL OF TROPICAL MEDICINE AND HYGIENE

Volume: Volume 104, Number 6

Publisher: AMER SOC TROP MED & HYGIENE | 2021-06-01, Pages 2275-2285

Type of Work: Article | Final Publisher PDF

Publisher DOI: 10.4269/ajtmh.20-0831

Permanent URL: https://pid.emory.edu/ark:/25593/vt7h9

Final published version: http://dx.doi.org/10.4269/ajtmh.20-0831

Copyright information:

© The American Society of Tropical Medicine and Hygiene

Accessed January 1, 2024 1:58 AM EST
Gut Microbiome Changes with Acute Diarrheal Disease in Urban Versus Rural Settings in Northern Ecuador

Maria J. Soto-Girón,1 Angela Peña-Gonzalez,1 Janet K. Hatt,2 Lorena Montero,3 Maritza Páez,3 Estefania Ortega,3 Shanon Smith,4 William Cevallos,5 Gabriel Trueba,3 Konstantinos T. Konstantinidis,1,2* and Karen Levy6*

1School of Biological Sciences, Georgia Institute of Technology, Atlanta, Georgia; 2School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia; 3Instituto de Microbiología, Universidad San Francisco de Quito, Quito, Ecuador; 4Rollins School of Public Health, Emory University, Atlanta, Georgia; 5Centro de Biomedicina, Universidad Central del Ecuador, Quito, Ecuador; 6School of Public Health, University of Washington, Seattle, Washington

Abstract. Previous studies have reported lower fecal bacterial diversity in urban populations compared with those living in rural settings. However, most of these studies compare geographically distant populations from different countries and even continents. The extent of differences in the gut microbiome in adjacent rural versus urban populations, and the role of such differences, if any, during enteric infections remain poorly understood. To provide new insights into these issues, we sampled the gut microbiome of young children with and without acute diarrheal disease (ADD) living in rural and urban areas in northern Ecuador. Shotgun metagenomic analyses of non-ADD samples revealed small but significant differences in the abundance of microbial taxa, including a greater abundance of Prevotella and a lower abundance of Bacteroides and Alistipes in rural populations. Greater and more significant shifts in taxon abundance, metabolic pathway abundance, and diversity were observed between ADD and non-ADD status when comparing urban to rural sites (Welch’s t-test, P < 0.05). Collectively our data show substantial functional, diversity, and taxonomic shifts in the gut microbiome of urban populations with ADD, supporting the idea that the microbiome of rural populations may be more resilient to ADD episodes.

INTRODUCTION

Urbanization has increased globally in recent years, with more than 53% of the total human population living predominantly in large cities.1 With the movement of populations from rural to urban areas comes changes in practices including eating behavior (diet rich in fat, simple sugars, and animal proteins), reduction in physical activity, increased use of antibiotics, and hygiene practices that are associated with an elevated prevalence of metabolic diseases including obesity, type 2 diabetes, and immunological diseases.2,3 These factors also impact the ecology, diversity, and functionality of the gut microbiome.4–8

Several investigations have compared the taxonomic profile of gut microbes between traditional agriculturalist societies (e.g., Malawi, Yanomami, Burkina Faso, and hunter-gatherers from Tanzania) and western populations (e.g., United States).2–7,9–10 The emerging picture from these previous studies is that rural and remote populations around the world harbor higher fecal bacterial diversity than urban-industrialized populations. Furthermore, it has been shown that microbiota diversity can be altered significantly, in just a few microbial generations, in rural human populations that undergo rapid urbanization or migration to urban settings, and some low-abundance members of the microbial community can become extinct.10–12 For instance, members of the genus Treponema have been found in the gut of rural traditional populations but not in the gut of urban-industrialized counterparts.6 On the other hand, urbanization has influenced the acquisition of specific bacterial taxa associated with a diet high in protein and animal fat.13

Many of the aforementioned studies showing the importance of geographic factors in determining gut microbiome composition are based on 16S recombinant RNA (rRNA) gene amplicon sequencing. However, this approach provides insufficient genetic resolution to capture intraspecific variation and whole-genome functional potential. In addition, most studies have been focused on differences in populations from distinct geographic regions, often without accounting for confounding factors such as different cultural practices or dietary preferences. A small number of recent studies has found lifestyle factors to be important for microbial community composition and functional metabolic properties along an urbanization gradient in the same geographic area (e.g., same country or region).9,14,15 However, most of these studies are not carried out using an epidemiological framework, so there is little known about the functional implications of these differences in gut microbiome composition in terms of health and disease consequences.

Changes in microbial composition and reduction in bacterial diversity can facilitate the colonization of the gut by pathogenic bacteria and, likely, the loss of keystone species, altering the homeostasis of the gut microbiota and producing a dysbiotic system10,21 with decreased resilience. Ecological theory generally suggests that high gut microbial diversity may confer increased resilience to perturbation and colonization by enteric pathogens.22,23 Thus, individuals in less-urbanized populations, with more diverse gut microbiomes, may have elevated resilience to enteric infections.

In our study, we explored the functional implications of differences in the gut microbial community composition and tested the hypothesis that microbial diversity confers increased resilience to perturbation in less-urbanized populations. Specifically, we examined metagenomic profiles of fecal samples from young children living in Quito (Ecuador’s capital city) and villages located in a rural area of the Esmeraldas Province of Ecuador, and characterized the value of this diversity for resilience, defined as the number of taxonomic or functional abundance shifts in the microbiome with acute diarrheal disease (ADD) compared with the non-ADD state.
MATERIAL AND METHODS

Study population. This study is part of the *E. coli en Zonas Urbanas y Rurales* (EcoZUR), a case-control study focused on diarrhea conducted in four sites along an urban–rural gradient in northern Ecuador (Figure 1A)\(^2\): 1) Quito (Ecuador’s capital) represents the urban area, with a population size of ~2.67 million; 2) Esmeraldas (capital of Esmeraldas Province), with a population size of ~162,000; 3) Borbón, a town located in the Esmeraldas Province with ~5,000 inhabitants; and 4) rural villages located along the Cayapas, Santiago, and Onzole Rivers with an estimated population between 10 and 500 inhabitants. Subjects from the villages generally presented low educational attainment levels and limited economic resources. More details about the region of study can be found in works by others.\(^2\)–\(^3\) The ages of the participants ranged between 0 months and 78 years and were categorized as follows for this sub-study: newborn, 0–6 months; babies, 7–12 months; young children, 13 months–3 years; children, 4–7 years; pre-adolescents, 8–17 years; adults, 18–74 years; and seniors, > 74 years.

The inclusion criteria comprised individuals visiting the clinic and recruitment centers with acute diarrhea (i.e., three or more loose stools in a 24-hour period [ADD samples]). Subjects from rural villages were enrolled during clinical visits onsite by Ministry of Health teams or at the Borbón Hospital. Non-ADD samples were site- and age-matched individuals visiting the same clinic for any other reason, who did not have symptoms of diarrhea or vomiting in the prior week. Both cases and controls were excluded if they had taken antibiotics in the prior week or if they had not lived in the study location for at least 6 months prior to presentation. Written informed consent was obtained from each participant and, for children and babies, was obtained from their parents and/or legal guardians at enrollment. This study was approved by the Institutional Review Board of Emory University and the Universidad San Francisco de Quito (USFQ). Participants were administered a survey to collect information about lifestyle and demographic factors including water consumption (source of drinking water, treatment), sanitation practices, contact with animals, recent travel, and other factors.

Sample collection. Fecal samples from participants were collected from April 2014–September 2015 and were stored in two cryo-conservation tubes and in a liquid nitrogen dewar until transferred to a –80°C freezer at the USFQ laboratory. Stool samples were cultured and tested for diarrheagenic *Escherichia coli* based on polymerase chain reaction screening for the presence of virulence genes, as described previously\(^2\): *aggR* for Enterogauggevive *E. coli,*\(^2\) *bfp* for typical enteropathogenic *E. coli,* *eaeA* for atypical enteropathogenic *E. coli,*\(^2\) *lt* and *sta* for enterotoxigenic *E. coli,*\(^2\) *ipaH* for enteroinvasive *E. coli* and Shigella,\(^6\) and *afa* for diffusely adherent *E. coli.*\(^3\) Positive pools for the *eaeA* gene were tested for *stx1* and *stx2* to detect enterohemorrhagic *E. coli.*

16S rRNA gene sequence analysis. DNA was extracted using the MoBIO PowerSoil DNA isolation kit (Qiagen, Germantown, MD) according to the Human Microbiome Project (HMP) protocol. The DNA concentration was estimated using
the Qubit 2.0 double-stranded DNA high-sensitivity assay (Invitrogen, Waltham, MA) Amplicon sequencing of the V4 region of the 16S rRNA gene was performed using primers 515F and 806R tagged with Illumina adapters P5 and P7.31 The tagged amplicons were sequenced with a MiSeq instrument and a 2 × 250-bp run.

Quality control and processing of raw paired-end reads were performed using DADA232 incorporated in QIIME2 v. 2017.9.35 The DADA2 denoise-paired plugin was used to trim low-quality regions of the sequences (less than Q30), remove chimeras, dereplicate sequences, and, finally, to produce an amplicon sequence variant table (hereinafter referred to as an operational taxonomic unit [OTU]) because of its analogy with the OTU table. To classify OTUs taxonomically, the QIIME2 q2-feature-classifier plugin and the Naive Bayes classifier were used together with the Greengenes13.8 99% OTUs database.34 The QIIME2 q2-diversity module was used to calculate alpha and beta diversity indexes based on a sampling depth of 8,000 reads/sample for all samples. This number of reads was used because coverage curves of randomly selected samples suggested that more than 99% of the community diversity was covered by our sequencing effort.

For alpha diversity, the number of observed OTUs and Faith’s phylogenetic diversity index were calculated. The Shannon diversity index with the Chao-Shen correction35 was calculated using the entropy package v. 1.2.136 available in R v. 3.3.1. For beta diversity, unweighted UniFrac and Bray-Curtis distances were calculated, and the distance matrices were the input for principal coordinate analyses (PCoAs). PCoA plots were visualized with EMPeror.37 Permutational multivariate analysis of principal coordinate analyses (PCoAs) were used to estimate the average coverage and diversity (similar to the Shannon index) for each sequenced library. Mash distances45 were calculated using a kmer = 25 and were visualized in a nonmetric multidimensional scaling plot using the ecodist46 package in R v. 3.3.1. Taxonomic classification of the short-read metagenomes was determined using Metaphlan2 with default parameters47 and the functional profile using HUMAnN2 with default parameters.48

Statistical analyses of taxonomic and functional profiles between samples for each group of subjects (i.e., rural versus urban) and with and without ADD were performed with STAMP v. 2.1.3 software.46 Welch’s t-test was used to compare relative abundances between the two groups. Correction for multiple comparisons was adjusted using the Benjamini-Hochberg FDR method (Q value).

**Recovery of genome populations in the metagenomes.** Assembled contigs longer than 1 Kb from each sample were binned into metagenome-assembled genomes (MAGs) using MaxBin2 with default parameters.50 Completeness and contamination of MAGs were estimated using CheckM v. 1.0.5 with the lineage_wf parameter.51 MAGs with > 85% completeness and < 8% contamination were selected for subsequent analyses. MAGs were annotated using Prokka v. 1.10 with default parameters,52 and predicted genes were mapped to the UniProt/SwissProt database using BLASTp v. 2.2.29+54 (minimum amino acid identity, ≥ 40%; query aligned length, ≥ 70% for a match). UniProt identifications were cross-referenced with gene ontology terms55,56 for assigning biological processes.

**Data availability.** All metagenomic data sets and 16S rRNA gene amplicon sequences used in our study are available in the National Center for Biotechnology Information (NCBI) repository under Bioproject no. PRJNA486009. MAGs can be found at http://enve-omics.ce.gatech.edu/data/ecozur_magsh.

**RESULTS**

**Cohort description and inclusion criteria.** Our study is part of the EcoZUR project, a case-control study of diarrhea conducted in four sites along an urban-rural gradient in northern Ecuador (Figure 1A).24 We selected subjects from the most urban and most rural sites to increase the potential for differences. Quito (Ecuador’s capital) represents the urban area, with a population size of ~2.7 million. Rural villages located along the Cayapas, Santiago, and Onzole Rivers, each with an estimated population between 10 and 500 inhabitants, represent the rural areas. Subjects from the villages generally presented low educational attainment levels and limited
economic resources. More details about the region of study are described earlier and can also be found in work by others. The inclusion criteria comprised individuals visiting clinics with acute diarrhea (i.e., more than three loose stools in a 24-hour period [ADD samples]). Controls (non-ADD samples) included age-matched individuals visiting the same clinic for any other reason, who did not have symptoms of diarrhea or vomiting in the prior week. Both cases and controls were excluded if they had taken antibiotics the previous week or if they had not lived in the study location for at least 6 months to avoid effects on the microbiome from migration.

We selected a subset of case samples from individuals symptomatic for ADD and a set of age-matched controls from the same site (Quito, 87; villages, 95) (Supplemental Table 1). The ages of the participants in the main EcoZUR study ranged between 0 months and 78 years. These samples were subjected to 16S rRNA gene amplicon sequencing. To reduce the impact of age as a covariate, and because diarrhea is predominantly a disease that affects young children, consistent with a higher number of samples from young children in our collection, we carried out shotgun metagenomic analysis on a randomly selected subset of samples from young children (1–6 years old). Samples for shotgun metagenomic analyses included 16 and 15 samples from non-ADD and ADD subjects in Quito (urban), respectively, and 17 and 15 samples from non-ADD and ADD subjects in the rural villages, respectively (Supplemental Table 2).

Geographic location effects on gut microbiota composition. Metrics of microbial richness and diversity based on 16S rRNA gene amplicons (Shannon index, phylogenetic diversity, and the number of observed OTUs) indicated minor, nonsignificant differences between urban and rural populations (Welch’s two-sample test, \( P > 0.05 \)) (Figure 1B). Nonetheless, when Bray-Curtis dissimilarity distances were plotted using PCoA, bacterial communities from non-ADD subjects living in urban areas segregated from the rural ones (PERMANOVA, \( P = 0.01 \)) (Figure 1C). Several samples were intermixed between the two groups, and we previously documented elevated risk of diarrhea and vomiting in the prior week. Both cases and controls were excluded if they had taken antibiotics the previous week or if they had not lived in the study location for at least 6 months to avoid effects on the microbiome from migration.

We selected a subset of case samples from individuals symptomatic for ADD and a set of age-matched controls from the same site (Quito, 87; villages, 95) (Supplemental Table 1). The ages of the participants in the main EcoZUR study ranged between 0 months and 78 years. These samples were subjected to 16S rRNA gene amplicon sequencing. To reduce the impact of age as a covariate, and because diarrhea is predominantly a disease that affects young children, consistent with a higher number of samples from young children in our collection, we carried out shotgun metagenomic analysis on a randomly selected subset of samples from young children (1–6 years old). Samples for shotgun metagenomic analyses included 16 and 15 samples from non-ADD and ADD subjects in Quito (urban), respectively, and 17 and 15 samples from non-ADD and ADD subjects in the rural villages, respectively (Supplemental Table 2).

Geographic location effects on gut microbiota composition. Metrics of microbial richness and diversity based on 16S rRNA gene amplicons (Shannon index, phylogenetic diversity, and the number of observed OTUs) indicated minor, nonsignificant differences between urban and rural populations (Welch’s two-sample test, \( P > 0.05 \)) (Figure 1B). Nonetheless, when Bray-Curtis dissimilarity distances were plotted using PCoA, bacterial communities from non-ADD subjects living in urban areas segregated from the rural ones (PERMANOVA, \( P = 0.01 \)) (Figure 1C). Several samples were intermixed between the two groups, and we previously documented elevated risk of diarrhea and diarrheagenic E. coli infections associated with travel to urban destinations. Therefore, the travel pattern of the corresponding human subjects, including the number of times reported to have traveled to the nearest town (Borbón), to a city (Esmeralda), and other communities in the past year, was evaluated to test whether traveling could explain the intermixing. No significant association between travel pattern and the overlap of the samples was found (Supplemental Table 3). Comparison between non-ADD subjects living in Ecuador (in urban and rural sites) and in the United States indicated that the bacterial communities were distinct (PERMANOVA, \( P = 0.01 \)) (Figure 1D) and that subjects living in Ecuador present greater microbial richness (average Chao1, 271.6 ± 61.3 versus 102 ± 36.9). A linear discriminant analysis effect size (LEfSe) identified the taxa that underlie this separation as greater proportions of Bacteroides, Clostridium, Coprococcus, and Faecalibacterium in U.S. subjects and greater proportions of Prevotella, Tepontia, Desulfovibrio, and Fusobacterium in Ecuadorian subjects (Supplemental Figure 1).

The influence of lifestyle and geographic variables on the fecal microbiota composition was evaluated with PERMANOVA of the 16S rRNA gene amplicon data. PERMANOVA showed that rural versus urban location (Adonis \( R^2 = 0.02, P = 0.001 \)), age (Adonis \( R^2 = 0.06, P = 0.001 \)), race (Adonis \( R^2 = 0.02, P = 0.027 \)), and reported water treatment (Adonis \( R^2 = 0.04, P = 0.012 \)) had a significant effect on the composition of gut microbiota (Table 1). Because this study was designed to focus on the effects of geography, rural location versus urban location was considered the main variable in the microbiome comparisons between participants with and without ADD from Quito and villages. Age was controlled for when we compared at the metagenomic level to reduce the impact of this variable as a covariate. We were unable to achieve the same for race, however, because of the low number of samples from mestizos individuals in the villages. Hence, the effect of race was controlled based on appropriate statistical methods. The remaining factors analyzed (delivery mode, education, house sanitation, water treatment type, gender; \( P > 0.05 \)) did not present significant correlations with microbial community structure.

To identify the specific microbial taxa responsible for the differences observed between rural and urban populations, we applied a multivariate association with a linear model in 16S rRNA gene amplicon data sets from non-ADD subjects, controlling for age and race (Supplemental Table 4). We

**Table 1**

<table>
<thead>
<tr>
<th>Variable</th>
<th>df</th>
<th>Sum of squares</th>
<th>Mean squares</th>
<th>F model</th>
<th>( R^2 )</th>
<th>P value (&lt; F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>1</td>
<td>1.82</td>
<td>1.82</td>
<td>4.98</td>
<td>0.02</td>
<td>0.001§</td>
</tr>
<tr>
<td>Age</td>
<td>5</td>
<td>4.12</td>
<td>0.82</td>
<td>2.25</td>
<td>0.06</td>
<td>0.001†</td>
</tr>
<tr>
<td>Race</td>
<td>4</td>
<td>1.82</td>
<td>0.45</td>
<td>1.24</td>
<td>0.02</td>
<td>0.02†</td>
</tr>
<tr>
<td>Delivery mode</td>
<td>3</td>
<td>1.31</td>
<td>0.43</td>
<td>1.0</td>
<td>0.02</td>
<td>0.071</td>
</tr>
<tr>
<td>Education</td>
<td>3</td>
<td>0.85</td>
<td>0.28</td>
<td>0.77</td>
<td>0.01</td>
<td>0.977</td>
</tr>
<tr>
<td>House sanitation</td>
<td>2</td>
<td>0.74</td>
<td>0.37</td>
<td>1.01</td>
<td>0.01</td>
<td>0.424</td>
</tr>
<tr>
<td>Water treatment type</td>
<td>2</td>
<td>0.94</td>
<td>0.47</td>
<td>1.29</td>
<td>0.01</td>
<td>0.059</td>
</tr>
<tr>
<td>Water treatment</td>
<td>1</td>
<td>0.62</td>
<td>0.62</td>
<td>1.71</td>
<td>0.008</td>
<td>0.012</td>
</tr>
<tr>
<td>Gender</td>
<td>1</td>
<td>0.29</td>
<td>0.29</td>
<td>0.81</td>
<td>0.004</td>
<td>0.789</td>
</tr>
<tr>
<td>Residuals</td>
<td>158</td>
<td>57.7</td>
<td>0.36</td>
<td>–</td>
<td>0.82</td>
<td>–</td>
</tr>
<tr>
<td>Total</td>
<td>180</td>
<td>70.32</td>
<td>–</td>
<td>1</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

\[^{\dagger}]=\text{degrees of freedom.}\]

\[^{\star}]=\text{R}^2\text{ values indicate the amount of variation attributed to each categorical parameter. PERMANOVA was conducted on the Bray-Curtis distance matrix using the ADONIS function in the vegan R package with 999 permutations of residuals.}\}

\[^{\times}]=\text{P}<0.001.\}
assessed non-ADD subjects only to understand gut microbiome conditions across the sites, without the influence of diarrhea symptoms. These results suggested that *Prevotella copri* (average abundance 37.53 ± 23.65% versus 19.57 ± 20.26%), and members from the *Comamonas* genus (average abundance 0.02 ± 0.04% versus 0.00%) and the *Elusimicrobiaceae* family (average abundance 1.10 ± 3.08% versus 0.05 ± 0.06%) were associated positively with rural settings, whereas members from the *Rikenellaceae* family (average abundance 1.99 ± 2.37% versus 0.30 ± 0.56%) were inversely associated with this location.

**OTU networks in rural versus urban microorganisms without the influence of diarrhea symptoms.** We explored the inter-microbial associations in the gut microbiota from non-ADD subjects living in rural and urban settings using OTU networks. This analysis estimates the inverse covariance of non-ADD subjects living in rural and urban settings using OTU networks. On the other hand, OTUs affiliated with *Oscillospira* presented the greatest number of edges (followed by *B. dorei* of the order of *S24-7* of the *Rikenellaceae* family) (average abundance 1.10 ± 3.08% versus 0.05 ± 0.06%) were associated positively with rural settings, whereas members from the *Rikenellaceae* family (average abundance 1.99 ± 2.37% versus 0.30 ± 0.56%) were inversely associated with this location.

**OTU networks in rural versus urban microorganisms without the influence of diarrhea symptoms.** We explored the inter-microbial associations in the gut microbiota from non-ADD subjects living in rural and urban settings using OTU networks. This analysis estimates the inverse covariance of non-ADD subjects living in rural and urban settings using OTU networks. On the other hand, OTUs affiliated with *Oscillospira* presented the greatest number of edges (followed by *B. dorei* of the order of *S24-7* of the *Rikenellaceae* family) (average abundance 1.10 ± 3.08% versus 0.05 ± 0.06%) were associated positively with rural settings, whereas members from the *Rikenellaceae* family (average abundance 1.99 ± 2.37% versus 0.30 ± 0.56%) were inversely associated with this location.

**Taxonomic affiliations of the 10 most connected OTUs differed between the two networks (Supplemental Table 6). In the rural network, *Bacteroides uniformis* showed the greatest number of connections (N = 12), followed by members of the family S24-7 of the order of *Bacteroidales* (N = 10), and *Oscillospira* (N = 10). On the other hand, OTUs affiliated with *Oscillospira* presented the greatest number of edges (N = 14), followed by *Bifidobacterium* (N = 10), and members of the *Erysipelotrichaceae* family (N = 10) in the urban network. The OTU classified as *Oscillospira* was the only highly connected OTU found in both groups. This taxon is a butyrate producer and is able to metabolize glucuronate, an animal-derived sugar, offering beneficial effects on human health.

**Shotgun metagenomic analysis of urban and rural microbiomes. Taxonomic differences.** To obtain a higher resolution of the microbiome structure associated with rural and urban lifestyles, non-ADD metagenomic data sets (16 from Quito and 17 from the villages) were compared using Mash similarity distances, showing that most of the metagenomics data sets from Quito clustered together whereas those from the rural villages were more spread out in the ordination plot (PERMANOVA, pseudo-F = 3.40, P < 0.05) (Supplemental Figure 2), consistent with the 16S rRNA gene-based results reported earlier.

**Shotgun metagenomic analysis of urban and rural microbiomes. Taxonomic differences.** To obtain a higher resolution of the microbiome structure associated with rural and urban lifestyles, non-ADD metagenomic data sets (16 from Quito and 17 from the villages) were compared using Mash similarity distances, showing that most of the metagenomics data sets from Quito clustered together whereas those from the rural villages were more spread out in the ordination plot (PERMANOVA, pseudo-F = 3.40, P < 0.05) (Supplemental Figure 2), consistent with the 16S rRNA gene-based results reported earlier.

**Inspection of taxonomic profiles of non-ADD subjects indicated the presence of taxa with differential abundance between rural and urban microorganisms (Figure 2A, B), despite a high overall inter-person variation. Relative abundance of *Prevotella* was significantly greater in the rural compared with the urban microbiomes (49.38 ± 28.62% versus 11.38 ± 22.10%, respectively; Tukey-Kramer post hoc test, P < 0.05). At the species level, the fraction of the *Prevotella* signal among the rural metagenomes was dominated by *P. copri* up to 100% of the total fraction in 8 of 17 samples) and *P. stercorae* (e.g., 93% in one sample). *Comamonas* and *Elusimicrobiaceae* members associated previously with rural samples based on 16S rRNA gene data did not show signatures of differential abundance based on the metagenomes. This pattern may be attributable to the fact that these taxa were mostly abundant in adult samples whereas the metagenomes were derived from young children.

In contrast, *Bacteroides*-affiliated populations were more abundant in urban compared with rural subjects (35.65 ± 27.67% versus 16.04 ± 18.74%, respectively; Tukey-Kramer post hoc test, P < 0.05). Of the 19 identified *Bacteroides* species, *B. dorei* was the most abundant (covering more than 93% of the total *Bacteroides* population in two samples), followed by *B. caccae* (more than 80% in two samples), and *B. vulgatus* (49% in one sample). *Alistipes* was also more abundant in urban compared with rural metagenomes (8.28 ± 10.79% versus 2.11 ± 4.00%, respectively; Tukey-Kramer post hoc test, P < 0.05). Of the eight identified *Alistipes* species, *A. shahii* was the most abundant (95% of the total *Alistipes* population in one sample), followed by *A. finegoldii* (100% in one sample), and *A. putredinis* (more than 70% in three samples).

Metagenome-assembled genome analysis showed a greater number of *Prevotella* MAGs in rural metagenomes than the urban ones (five MAGs versus one MAG, respectively), whereas *Alistipes* MAGs were only recovered from urban metagenomes (Supplemental Table 7). Taxonomic assignment of *Prevotella* and *Alistipes* MAGs was confirmed by estimating the average nucleotide identity (ANI) between the MAG and reference genomes from the NCBI (NCBI_Prok) using MiGA. This analysis showed that most of the *Prevotella* MAGs (67%) represented uncharacterized species, defined at the 95% ANI level, relative to previously described species (Supplemental Table 7).

**Analysis of the metabolic potential of *Prevotella* and *Alistipes* MAGs revealed 50 biological processes (Supplemental Figure 3A) and 71 enzymatic reactions to be differentially abundant (Supplemental Figure 3B) (Welch’s t-test, P < 0.05 with Benjamini-Hochberg FDR correction). Functional annotation of protein-coding genes indicated that *Prevotella* MAGs harbored a greater number of genes related to amino acids (arginine, methionine), cofactors (pyridoxal phosphate), and nucleotide biosynthesis (adenine, nicotinamide adenine di-nucleotide, adenosine monophosphate, purine salvage), and metabolic processes (arginine and guanine catabolism) than *Alistipes* MAGs. Several of these pathways have been associated with beneficial effects in the host, such as synthesis of essential amino acids, energy production, protein synthesis, and intestinal barrier function, in addition to acting as precursors for several microbial metabolic pathways. Moreover, some of the enzymes encoded in the *Prevotella* MAGs participate in polysaccharide metabolism (e.g., glycoside hydrolases, glycans transferases) and thus are likely associated with diet. *Prevotella* has been associated previously with a diet high in fiber, carbohydrates, vegetables, and egg food items. On the other hand, *Alistipes* MAGs showed a greater number of predicted pathways and enzymes associated with carbohydrate (ribose, pentose phosphate), phospholipid (cardiolipin), and secondary metabolism; cell response to starvation; and cellular protein modification processes. *Alistipes* has been associated previously with a meat-based diet.

**Functional gene differences.** Functional annotation of metagenomic reads from all samples was performed using
HUMAN2 and the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Figure 3C). The predicted microbial functions highlighted differences in 35 KEGG pathways associated with non-ADD samples between the rural and urban groups. In particular, the 35 pathways were associated with biosynthesis pathways, including nucleotides, amino acids, cofactors, and secondary metabolites as well as glycolysis and carbohydrate metabolism pathways. Among these, pathways related to ribonucleotide biosynthesis (average relative abundance in the metagenome 0.0012 ± 0.0004 versus 0.0008 ± 0.0002), lysine (0.0008 ± 0.0003 versus 0.0006 ± 0.0002), methionine (0.0008 ± 0.0003 versus 0.0005 ± 0.0002), and methylerythritol phosphate (0.0005 ± 0.0001 versus 0.0008 ± 0.0003) pathways showed significantly greater abundance in the microbiota of rural compared with urban samples. Some of these pathways may play key roles in maintaining intestinal homeostasis. For instance, chorismate is a precursor for many bacterial metabolic pathways,69 and amino acid metabolism by bacteria is thought to be an important modulator of diverse physiological processes.65 Correlation analysis between bacterial taxonomic and functional richness indicated that, in rural samples, the number of predicted genes correlates significantly with the number of OTUs present in the metagenomes (Pearson’s r = 0.71, P < 0.01) (Supplemental Figure 4).

**Figure 2.** Differences in taxonomic and metabolic features in the non-acute diarrheal disease shotgun metagenomes from urban versus rural young children: Taxonomic and functional analyses of a subset of the total samples from young children that were subjected to shotgun metagenomic sequencing: (A) Principal component (PC) analysis of relative abundances of microbial members at the family level between the two groups of samples. (B) Bar plots showing the proportion of sequences assigned to each differentially abundant taxon, with the median (black central line) and the mean (white star). Small panels at the bottom indicate the mean proportion differences and CIs at 95% of *Prevotella copri* and *Prevotella stercorea* in rural and urban microbiomes (Tukey-Kramer post hoc test, P < 0.05). (C) Heat map showing the relative abundance of 35 Kyoto Encyclopedia of Genes and Genomes pathways that were significantly differentially abundant between the two groups (corrected P value < 0.05 after multiple comparisons). This figure appears in color at www.ajtmh.org.

**Shifts in the gut microbiome with ADD.** We profiled the gut microbiota with diarrhea in subjects from rural and urban settings to test for differences in the microbiome with ADD and the potential role of the taxonomic and functional differences identified earlier from the non-ADD state comparisons. 16S rRNA gene sequence analysis showed that samples from urban subjects presented a lower number of observed OTUs with ADD than those from rural subjects (average number: 99 versus 128, respectively; Welch’s t-test, P < 0.05) despite the comparable number of OTUs in the non-ADD state (described earlier). A significant decrease in diversity with diarrhea in urban subjects when compared with non-ADD samples was also observed (Shannon index with the Chao-Shen correction average, 20 ± 13 versus 27 ± 16; Welch’s t-test, P < 0.05), but not in rural subjects (Supplemental Figure 5A).

OTU networks from rural and urban subjects indicated a change in the network topology and connectivity patterns with ADD compared with the non-ADD state (Supplemental Figure 5B and Supplemental Table 4). Specifically, the urban network showed a greater reduction, by 30% or more, relative to the non-ADD network in both the number of nodes (OTUs) and connections with diarrhea. Furthermore, many of the OTUs with the greatest number of connections appeared to be lost with ADD (Supplemental Figure 5B). On the other hand,
the rural network presented 6% fewer nodes and 21% fewer connections with ADD than the one without ADD. This network also presented a reduction of the most connected OTUs, but the effect appeared to be less pronounced than the one in the urban network.

Different pathogenic agents can cause diarrhea, including infectious and noninfectious agents that produce different alterations in the intestinal epithelium. Although the causative agent of diarrhea of our ADD samples is unknown, we attempted to confirm the presence of pathogenic *E. coli* in the gut metagenomes and its possible association with diarrhea/observed phenotype using our recently proposed methodology and Supplemental Experimental Procedures. This methodology combines four independent lines of evidence to identify the probably causative agent—that is, the abundance estimation of suspected pathogen (e.g., pathogenic *E. coli*) in the metagenomes, the presence of virulence factors, the degree of clonality (e.g., pathogenic populations are typically more clonal than their commensal counterparts), and the frequency of clonal complexes associated with a disease state. Results of these analyses indicated that five ADD samples from urban subjects and two from rural subjects may harbor pathogenic *E. coli* (Supplemental Table 7).

Metagenomic comparison after removing the *E. coli* signal in ADD samples showed similar results from those comparing all ADD samples at the taxonomic and functional profiles (Supplemental Results). Therefore, we proceeded with our analysis using all samples. Significant shifts in metagenomic abundance with diarrhea were observed in members of the *Desulfovibrionaceae* family in rural samples, whereas members of the *Bacteroidaceae*, *Porphyromonadaceae*, and *Ruminococcaceae* families presented a reduction in abundance in the urban metagenomes (Tukey-Kramer post hoc test, *P* < 0.05) (Figure 3A). Changes in abundance in these taxa have been associated previously with infectious diarrhea. In addition, five obligate anaerobes were significantly depleted in metagenomes from urban samples with ADD (unclassified *Subdoligranulum*, *Desulfovibrio piger*, *Roseburia hominis*, *Parabacteroides distasonis*, *Ruminococcus obeum*; Tukey-Kramer post hoc test, *P* < 0.05), which has also been observed previously.

When comparing the metabolic diversity encoded by metagenomic reads, urban but not rural subjects showed significantly lower diversity in ADD compared with non-ADD (Welch’s *t*-test, *P* < 0.05) (Figure 3B). In addition, a greater number of predicted metabolic pathways changed in abundance in urban versus rural samples in the ADD state (28 versus three pathways, respectively; Figure 4A). The bulk of these pathways were involved in the biosynthesis of various cofactors (coenzyme A, pyridoxal phosphate, folate, N10-formyl-tetrahydrofolate) and amino acids (ornithine, arginine, and polyamine) as well as carbohydrate metabolism (mannan degradation, gluconeogenesis, glycolysis). A decreased relative abundance of some of these pathways (e.g., amino acids) was also observed in the rural network.

![Figure 3](https://www.ajtmh.org/)
Acid biosynthesis, glycolysis and pentose phosphate pathway has been reported previously in patients with Clostridium difficile infection. On the other hand, the metabolic pathways with a depleted abundance in diarrhea in rural populations were involved in sulfur oxidation, urea cycle, and L-isoleucine biosynthesis (Figure 4A). These pathways have been associated with an elevated demand for energy production during mucosal inflammation and diarrhea affecting the urea cycle and amino acid levels in the colonic mucosal tissue.

Taxonomic annotations of KEGG pathways with significant shifts with ADD were used to identify those microbial members that were involved in the metabolic changes with diarrhea (Figure 4B). This analysis showed a different taxonomic profile between rural and urban metagenomes associated with the functional pathways altered significantly with ADD. For instance, Ruminococcus bromii, Coprococcus sp. ART55.1, and Treponema succinifaciens were involved in the reduction of abundance observed in sulfur oxidation, urea cycle, and L-isoleucine biosynthesis pathways in rural samples. On the other hand, Faecalibacterium prausnitzii, Alistipes shahii, and different Bacteroides species—including B. ovatus, B. caccae, and B. fragilis—were among the taxa that participated in the reduction of metabolic pathways with ADD in urban subjects.

**DISCUSSION**

In our study we found significant differences between rural and urban microbiomes of young children with ADD in northern Ecuador, most likely associated with different taxa and metabolic pathways present in the non-ADD state between rural and urban subjects (Figure 3). Differences in lifestyle factors in rural versus urban areas such as dietary habits, social status, economic development, antibiotic accessibility, agricultural activities, and exposure to livestock most likely influence gut microbiome changes during ADD. Moreover, the significant correlation between richness of functional and taxonomic profiles found in rural but not in urban samples could underlie, at least in part, the smaller disturbance of the ADD microbiome observed in rural compared with urban children (Supplemental Figure 4). These taxonomic and functional differences might be important for the resilience of...
the microbiome to pathogen infections, although future work is required to elucidate the underlying mechanisms for the results observed.

We noted that the number of KEGG metabolic pathways and taxon abundance with significant shifts was greater in ADD urban samples in comparison with the rural ones (Figures 3A and 4A). However, the effects of ADD seem to be variable among samples (large standard deviation), presumably reflecting high inter-individual variation to ADD. Longitudinal sampling of a cohort of subjects (which we currently have under way) will test whether the differences in microbiome composition in the non-ADD state between rural and urban microbiomes play a significant functional role in modulating responses to enteric infections and ADD.

The high individual variability observed in ADD may be associated with distinct extrinsic (causative agent of diarrhea, geographic factors) and intrinsic (microbial diversity patterns, functional capacities, community structure, host genetics, host immune status) factors. The causative agents of diarrhea include a broad spectrum of microbial (e.g., bacteria, viruses, and protozoa parasites) and noninfectious agents (e.g., food indigestion, environmental exposures, inflammatory diseases).12 Regardless of the exact extrinsic and intrinsic factors underlying the high inter-subject variability, our results indicate collectively significant qualitative differences observed in the gut microbiome to ADD between urban and rural subjects that warrants further investigation. Consistent with these interpretations, when we identified and removed the ADD cases that were, most likely, caused by *E. coli* pathotypes (based on in situ abundance, level of clonality, and presence of virulence factors; Supplemental Experimental Procedures and Supplemental Results), our major conclusions did not change (Supplemental Figure 6).

Analysis of OTU networks of the gut microbiota indicated that rural populations presented a greater number of OTUs and more connections among OTUs when compared with the urban ones (Supplemental Table 5). A denser network may be associated presumably with more connections among distinct bacterial species, modulating stability and community assembly in the gut during perturbations. In this case, we observed that the ADD urban network showed shifts in more connections and nodes (e.g., loss of connections) than the rural one, indicating possibly a more altered and unstable community with a lower resilience to infection/diarrhea (Supplemental Table 5). This was consistent with the greater number of taxonomic and metabolic shifts identified in urban samples, which indicate a more resilient microbiome for rural subjects.

These observed differences in taxa abundance between the two groups in the non-ADD state might be attributable, at least in part, multiple (unmeasured) variables, including human contact with other communities and/or cities, exposure to animals/livestock, water quality, socioeconomic status, and race. For instance, we observed an unequal distribution of participants by race in the urban setting versus the rural setting, in which mestizos dominated the Quito group whereas Afro-Ecuadorians dominated the rural one (Supplementary Table 1). Unfortunately, because of the racial makeup of these Ecuadorian populations, it is not feasible to recruit people of African descent only in the Quito population, or people of mixed race only in the rural study site.

Although dietary preferences were not recorded in this study, our previous work in this region has found a diet based on fish, shrimp, plantain, rice, and coconut in rural villages.13 Subjects from the villages also use medicinal plants.81 On the other hand, a typical daily dish in Quito is composed of a high percentage of carbohydrates, proteins, and fats, and less than 15% of vegetables. Pork is one of the main ingredients in Ecuadorian Andean dishes.82 Moreover, it has been reported previously that housing density, road access, social connectedness, and food-sharing, among other factors, influence pathogen prevalence and transmission in a community.26,83 Presumably, all these factors combined drive the differences we observed between rural and urban subjects in the non-ADD state, although the relative importance of each factor should be investigated in future studies.

The microbial community structure determined in the sampled subjects from the rural villages resembled those reported previously for rural populations, including Yanomami (Venezuela), Malawian (Amazon), Hadza (Tanzania), and Mat-ses (Peru).5,6,9-11,13,65,84 In particular, the fecal microbiota from rural subjects showed a greater abundance of *Prevotella*, whereas *Bacteroides* and *Alistipes* presented a greater fraction of the total microbial community in the urban ones (Figure 2B). Most *Prevotella* and *Alistipes* MAGs recovered from the metagenomes represent novel diversity because they could not be assigned to known bacterial species, suggesting they may represent previously uncharacterized species (Supplemental Table 7). Stagaman et al.19 reported a similar taxonomic profile (e.g., abundance of *Prevotella* correlated negatively with house modernity, and an abundance of *Bacteroides* correlated positively) along a gradient of economic development in five villages close to the Cordillera de Cucutí in southeastern Ecuador. Together, these findings and those of our study in northern Ecuador indicate there might be some universal patterns accompanying the phenomenon of urbanization in Ecuador and elsewhere. Nonetheless, despite the presence of “western-type” taxa in subjects from Quito, comparisons to the gut microbiota from U.S. subjects revealed a clear segregation between these two locations (Figure 1D), suggesting that the gut microbiota of subjects living in Ecuador’s capital have not acquired the profile of a high-income industrialized country.

Collectively, our data reveal significant compositional differences in the gut microbiome of young children in rural and urban areas in northern Ecuador. Rural populations appeared to present smaller alterations to ADD in comparison with urban ones, potentially driven by differences in microbial composition, functional diversity, microbe–microbe associations, and stability. Future studies with larger and more balanced (in terms of race cohorts) longitudinal sampling (as opposed to the cross-sectional approach used here) to reduce the variability observed among participants, and with ADD samples with known etiological agents will help examine further the extent to which urbanization/lifestyle modulate physiological and microbial changes during enteric infections and ADD, especially in critical populations such as young children from developing countries, where ADD is a major cause of mortality.

Received July 9, 2020. Accepted for publication February 15, 2021.
Published online April 19, 2021.

Note: Supplemental information, tables, and figures appear at www.ajtmh.org.

Acknowledgment: We thank all members from Dr. Trueba’s lab for their help in collecting samples and extracting DNA. We are grateful to all volunteers who participated in this study.
Financial support: This study was supported by the National Institute for Allergy and Infectious Diseases (grant nos. K01AI103544 and R01AI137679) at the U.S. National Institutes of Health. M. S.- G. was supported by a doctoral scholarship from Colciencias, Colombian government.

Disclaimer: The content in this article is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Authors’ addresses: Maria J. Soto-Girón and Angela Peña-González, School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, E-mails: dendroapsis@gmail.com and angela.viviana.penal@gmail.com. Janet K. Hatt and Konstantinos T. Konstantinidis, School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA, E-mails: janet.hatt@ce.gatech.edu and kostas.konstantinidis@gatech.edu. Lorena Montero, Maritza Páez, Estefanía Ortega, and Gabriel Trueba, Instituto de Microbiología, Universidad San Francisco de Quito, Quito, Ecuador, E-mails: lpmontero@gmail.com, mpazelllena86@gmail.com, estefania1058@gmail.com, and gtrueba@usfq.edu.ec. Sharon Smith, Rollins School of Public Health, Emory University, Atlanta, GA, E-mail: shanon.smith@emory.edu. Karen Levy, School of Public Health, University of Washington, Seattle, WA, E-mail: klevyx@uw.edu. William Cevallos, Centro de Biomedicina, Universidad Central del Ecuador, Quito, Ecuador, E-mail: wcevallos@usfq.edu.ec.

REFERENCES
