



Published in final edited form as:

Sci Adv. 2015 April 3; 1(3): . doi:10.1126/sciadv.1500183.

The microbiome of uncontacted Amerindians

Jose C. Clemente^{1,2,*}, Erica C. Pehrsson^{3,*}, Martin J. Blaser^{4,5}, Kuldip Sandhu^{5,†}, Zhan Gao⁵, Bin Wang³, Magda Magris⁶, Glida Hidalgo⁶, Monica Contreras⁷, Óscar Noya-Alarcón⁶, Orlana Lander⁸, Jeremy McDonald⁹, Mike Cox⁹, Jens Walter^{10,‡}, Phaik Lyn Oh¹⁰, Jean F. Ruiz¹¹, Selena Rodriguez¹¹, Nan Shen¹, Se Jin Song¹², Jessica Metcalf¹², Rob Knight^{12,13,§}, Gautam Dantas^{3,14}, and M. Gloria Dominguez-Bello^{5,7,11,¶}

¹Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

²Immunology Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

³Center for Genome Sciences & Systems Biology, Washington University School of Medicine, St. Louis, MO 63108, USA

⁴Laboratory Service, VA Medical Center, New York, NY 10010, USA

⁵New York University School of Medicine, New York, NY 10016, USA

⁶Amazonic Center for Research and Control of Tropical Diseases (CAICET), Puerto Ayacucho 7101, Venezuela

⁷Venezuelan Institute for Scientific Research, Caracas 1020-A, Venezuela

⁸Sección de Ecología Parasitaria, Instituto de Medicina Tropical, Universidad Central de Venezuela, Caracas 1051, Venezuela

⁹Anaerobe Systems, Morgan Hill, CA 95037, USA

¹⁰Department of Food Science and Technology, University of Nebraska, Lincoln, NE 68583, USA

exclusive licensee American Association for the Advancement of Science. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

[¶]Corresponding author. maria.dominguez-bello@nyumc.org.

*These authors contributed equally to this work.

[†]Deceased.

[‡]Present address: Department of Biological Sciences and of Agricultural, Food, and Nutritional Science, University of Alberta, Alberta, Edmonton T6G 2R3, Canada.

[§]Present address: Departments of Pediatrics and Computer Science & Engineering, University of California at San Diego, La Jolla, CA 92093, USA.

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/1/3/e1500183/DC1>

Author contributions: M.M. and M.G.D.-B. planned the study; O.N.-A., M.M., G.H., M. Contreras, O.L., and M.G.D.-B. implemented the sample collection and preservation; Z.G., K.S., J. McDonald, M. Cox, S.R., J.F.R., J.W., P.L.O., and M.G.D.-B. characterized bacterial isolates; E.C.P., B.W., and G.D. characterized the resistome; R.K., Z.G., K.S., J. McDonald, M.G.D.-B., J.W., P.L.O., S.R., B.W., E.C.P., and G.D. generated data; J.C.C., R.K., M.J.B., E.C.P., G.D., N.S., S.J.S., J. Metcalf, and M.G.D.-B. performed analyses and interpreted results; J.C.C., E.C.P., G.D., M.J.B., R.K., and M.G.D.-B. wrote the paper.

Competing interests: The authors declare that they have no competing financial interests.

Data and materials availability: 16S rRNA sequence data has been deposited to EBI with IDs ERA387449 and ERP008799. Assembled functional metagenomic contigs have been deposited to GenBank with IDs KJ910904–KJ910951, KJ910953–KJ910954, KJ910958–KJ910959, KJ910961–KJ910973, KJ910975–KJ910989, KJ910991–KJ911003, KJ911005–KJ911006, KJ911008–KJ911009, KJ911011, and KP063058, and raw shotgun metagenomic reads to SRA with ID SRP049631.

¹¹Department of Biology, University of Puerto Rico, Rio Piedras 00931, Puerto Rico

¹²Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309, USA

¹³Howard Hughes Medical Institute, University of Colorado, Boulder, CO 80309, USA

¹⁴Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO 63110, USA

Abstract

Most studies of the human microbiome have focused on westernized people with life-style practices that decrease microbial survival and transmission, or on traditional societies that are currently in transition to westernization. We characterize the fecal, oral, and skin bacterial microbiome and resistome of members of an isolated Yanomami Amerindian village with no documented previous contact with Western people. These Yanomami harbor a microbiome with the highest diversity of bacteria and genetic functions ever reported in a human group. Despite their isolation, presumably for >11,000 years since their ancestors arrived in South America, and no known exposure to antibiotics, they harbor bacteria that carry functional antibiotic resistance (AR) genes, including those that confer resistance to synthetic antibiotics and are syntenic with mobilization elements. These results suggest that westernization significantly affects human microbiome diversity and that functional AR genes appear to be a feature of the human microbiome even in the absence of exposure to commercial antibiotics. AR genes are likely poised for mobilization and enrichment upon exposure to pharmacological levels of antibiotics. Our findings emphasize the need for extensive characterization of the function of the microbiome and resistome in remote nonwesternized populations before globalization of modern practices affects potentially beneficial bacteria harbored in the human body.

Introduction

Host-microbial interactions are important determinants of host physiology, including immune responses, metabolic homeostasis, and behavior (1–4). Microbiota transplantation can transfer phenotypes such as nutritional status from donor to recipient (5, 6), indicating that altered microbial communities can therefore cause, as well as result from, altered physiological states. Despite increasing evidence that the microbiome has important roles in human health, we do not yet know the extent to which the human microbiome has changed during the adoption of life-styles associated with westernization. Here, we described the microbiome from Yanomami subjects in the Amazon with no previous report of contact with non-Yanomami. The Yanomami were originally mountain people who were first contacted in the mid-1960s, and who continue to live seminomadic hunter-gatherer life-styles in the Amazon jungle. In Venezuela, they inhabit a region protected from development, and many still inhabit uncharted villages in the vast mountainous Yanomami territory (7). These remote populations of hunter-gatherers have life-styles similar to those of our human ancestors, and are unexposed to modern practices known to exert antimicrobial effects.

For instance, pharmacologic-dose antibiotic exposure is nearly ubiquitous worldwide, and all human microbiota studied to date, remote or industrialized, harbor diverse antibiotic

resistance (AR) genes (8, 9). Although AR genes have been computationally predicted in ancient oral microbiota (10), functional resistance from the preantibiotic era remains largely uncharacterized. These subjects from an uncontacted community therefore represent a unique proxy for the preantibiotic era human resistome. Here, we characterized the microbiome and resistome of these subjects, and compared them to those of other non-isolated populations. Samples of the oral cavity ($n = 28$), forearm skin ($n = 28$), and feces ($n = 12$) were obtained from 34 of the 54 villagers (table S1) at the time of the first medical expedition to an isolated, previously uncharted village in the High Orinoco state of Venezuela. The age of the subjects was between 4 and 50 years, as estimated by Yanomami health workers.

Results

Unprecedented high bacterial and functional diversity in the microbiome of uncontacted Amerindians

We sequenced the V4 region of the 16S *rRNA* gene from fecal, oral, and skin samples and compared these data with those from previous studies (11, 12). The microbiome of the uncontacted Amerindians exhibited the highest diversity ever reported in any human group. Diversity in feces and skin, but not in the mouth, was significantly higher in Yanomami than in U.S. subjects (Fig. 1, A, E, and I). Noticeably, fecal diversity was even higher than in the semitranscultured Guahibo Amerindians and Malawians [analysis of variance (ANOVA) and Tukey's honest significant difference (HSD) test, $P < 0.001$; Fig. 1A]. Principal coordinate analysis (PCoA) of UniFrac distances reveals that bacterial communities from each body site segregate westernized people from agrarian and hunter-gatherer groups (Fig. 1, B, F, and J, and fig. S1). Accumulation curves of fecal operational taxonomic units (OTUs) and phylogenetic species richness (γ diversity) confirmed that the Yanomami harbor significantly higher diversity than other populations (unpaired *t* test with Bonferroni correction, $P < 0.0001$; fig. S2A) and, at the same time, have the lowest microbiome variability (fig. S2B). Within-population distances in the Yanomami were lower than those in the semi-accultured Guahibo Amerindians (fig. S2C). Prevalence/abundance curves further confirm that the Yanomami share a higher proportion of their fecal microbiota despite their higher diversity (Fig. 1D). These results were consistent with those obtained sequencing the V2 region of the 16S *rRNA* gene of the same set of samples extracted, amplified, and sequenced together with samples from U.S. subjects (fig. S3).

On the basis of taxa detectable at the obtained sequencing depth, the Yanomami fecal microbiota was characterized by high *Prevotella* and low *Bacteroides*, contrary to what was observed in U.S. subjects and similar to that in Guahibo Amerindians, Malawians (fig. S4), and African hunter-gatherers (13). Analyses of discriminative bacteria in relation to U.S., Malawian, and Guahibo populations (14) indicated that Yanomami had higher intestinal Bacteroidales S24-7 (a poorly characterized family close to the *Parabacteroides*), Mollicutes and Verrucomicrobia, members of the families Aeromonadaceae, Oxalobacteraceae, and Methanomassiliococcaceae, and genera *Phascolarctobacterium*, *Desulfovibrio*, *Helicobacter*, *Spirochaeta*, and *Prevotella* (Fig. 1C). Although diet plays a significant role in shaping the microbiome, factors other than nutrition also influence

microbiota composition. For example, *Helicobacter* was absent in the feces of U.S. subjects, and intestinal *Oxalobacter* and *Desulfovibrio* were markedly increased in the Yanomami in relation to U.S. subjects (fig. S5).

In contrast to the fecal microbiome, the Yanomami had similar levels of oral bacterial diversity to those found in U.S. subjects (Fig. 1E) despite a distinct composition (Fig. 1F), and higher prevalence of oral taxa (Fig. 1H). Although both populations were dominated by *Streptococcus*, the Yanomami had higher proportions of several oral *Prevotella*, *Fusobacterium*, and *Gemella*, and lower *Rothia*, *Stenotrophomonas*, *Acinetobacter*, or *Pseudomonas* (Fig. 1G and fig. S4). The Yanomami cutaneous microbiota had significantly higher diversity (Fig. 1I) and clustered apart from that of U.S. subjects (Fig. 1J). No single taxon dominated the Yanomami skin microbiota, in contrast to the low skin bacterial diversity in U.S. subjects and high relative proportions of *Staphylococcus*, *Corynebacterium*, *Neisseriaceae*, and *Propionibacterium* (Fig. 1K and fig. S4). Some bacterial taxa previously reported as environmental, such as *Knoellia* or *Solibacteraceae* (15–17), appeared to be enriched in the skin microbiota of the Yanomami. As with fecal and oral microbiota, skin bacterial prevalence was also higher in the Yanomami (Fig. 1L).

Metagenomic predictions using PICRUSt (18) showed distinct microbiome functional profiles for different human populations (Fig. 2, B, F, and J), with the Yanomami having higher fecal and skin (Fig. 2, A and I), but not oral (Fig. 2E), functional diversity, higher gene prevalence (Fig. 2, D, H, and L), and lower intragroup variation (fig. S2D). Pathways involving amino acid metabolism, glycosyltransferases, and biosynthesis of lipopolysaccharides, terpenoid-quinones, and vitamins [including riboflavin, previously reported to be increased in non-Western populations (11)] were found to be overrepresented in the Yanomami (Fig. 2C). Their oral mucosa also had significant overrepresentation of pathways of fatty acid biosynthesis, glycosyltransferases, and methane metabolism (mostly derived from members of the Methylocystaceae) (Fig. 2G). Their skin was enriched in organic acids, amino acids, vitamins, and methane (originating from *Methylobacter*, *Methylosinus*, and Methylocystaceae) pathways (Fig. 2K). Validation of the predicted bacterial functions by comparing against the sequenced metagenome of the same samples and using a previously reported metagenome from Guahibo Amerindians (11) confirmed the robustness of our results (see Materials and Methods).

To provide an archive to study human microbial evolution, we isolated 1097 bacterial strains from Yanomami fecal or oral samples. Characterization of 24 Yanomami *Escherichia coli* strains using multilocus sequence typing (MLST) based on seven housekeeping genes (19) showed that they differ substantially from strains circulating in Western societies, and that none of the isolates belonged to phylogenetic group B2 common in Western populations (fig. S6).

Presence of AR genes in the microbiome of uncontacted Amerindians

Phenotypic resistance was tested in 131 Amerindian *E. coli* strains isolated from 11 fecal specimens. We found that all *E. coli* isolates were sensitive to all 23 antibiotics tested. However, functional libraries created from the genomic DNA of 38 of these *E. coli* isolates, including the 24 typed strains (table S12), yielded AR genes targeted against eight

antibiotics, including four antibiotics from within the phenotypic screen (table S3). Overexpression of chromosomally encoded *E. coli* genes was likely responsible for most of the resistance (89% have homologs in *E. coli* K-12 MG1655; table S4).

For a culture-independent interrogation of the commensal Amerindian resistome, we created shotgun metagenomic libraries from fecal and oral microbiota of four subjects (table S5) (8, 20) after multiple displacement amplification (MDA) and screened them against 15 antibiotics (table S6). We identified 28 functional AR genes in the microbiota of these people naïve to anthropogenic antibiotics (table S4), including genes resistant to semisynthetic and synthetic antibiotics, to which the villagers and their surroundings are highly unlikely to have ever been exposed. These genes include an extended-spectrum *cblA* β -lactamase that confers resistance to five later-generation antibiotics, including a fourth-generation cephalosporin and a synthetic monobactam, and penicillin-binding proteins (PBPs) resistant to the third-generation cephalosporin ceftazidime (Fig. 3A). The six PBPs have 81 to 100% amino acid identity to PBPs from *Neisseria* and *Kingella*, oral commensals in both Western and Amerindian hosts (21).

Most of the AR genes had close homologs in human microbiota from industrialized societies. Seventy-nine percent of the AR genes aligned to the Human Microbiome Project (HMP) assemblies with 95% nucleotide identity, and all genes that aligned to HMP (87% of fecal genes) also aligned to MetaHIT assemblies (Fig. 3B). Only one gene (F3_TE_1), a ribosomal protection protein (RPP) encoding resistance to tetracycline, did not align to HMP, MetaHIT, or any protein in National Center for Biotechnology Information (NCBI) nr (at >70% amino acid identity) and may be exclusive to this population. Thus, despite different antibiotic exposures, the microbiota from antibiotic-naïve and industrialized people share a common resistome.

Whole metagenome shotgun sequencing of fecal and oral specimens from the Amerindian village was performed to quantify the abundance of the functionally selected AR genes from the Amerindian microbiota and from age- and gender-matched individuals from an industrialized setting (Puerto Rico). Genes from all resistance classes except RPPs were restricted to their body site of origin, either the fecal or oral microbiota (Fig. 4), perhaps due to tight linkage with a microbial host specific to that habitat. Antibiotic-resistant PBPs and transporters functionally selected from Puerto Rican metagenomes were no less abundant in the Amerindian metagenomes (Wilcoxon rank sum test, Bonferroni correction, $P > 0.05$), and 72% share 95% amino acid identity with commensal *Neisseria* and *Kingella* species. However, 14% of the Amerindian AR genes are syntenic with mobilization elements, including plasmids and transposases, highlighting the risk for horizontal gene transfer in response to antibiotic selection (see the Supplementary Materials) (22). This is likely an underestimation of the mobilization risk, because many of the AR genes not found to be syntenic with mobilization elements in the Amerindian context are associated with such elements in industrialized settings and are widespread in commensals and pathogens (Fig. 3, C and D).

Discussion

The characterization of an isolated group of Yanomami hunter-gatherers here reported represents a unique opportunity to understand a human microbiome and resistome minimally exposed to modern practices, which can exert significant antimicrobial effects with lasting consequences in the host (23, 24). The Yanomami here described exhibit unprecedented levels of bacterial diversity. Despite the small population of this village and the limited sample size of our study, they had significantly higher levels of fecal and skin diversity, both bacterial and functional, than any other human population reported so far. In contrast with the large infrequent meals characteristic of Western diets, frequent meals and food seasonality in the Yanomami would, according to disturbance theory (25, 26), lead to oscillating microbial populations depending on resource availability. The microbiome diversity is more evenly shared among the Yanomami across all body sites than in other populations, a phenomenon that could partially be due to cohabitation (27). However, Guahibo Amerindians living in two different Amazonian villages had less shared fecal taxa and higher within-village distances (fig. S2C). Together, these results suggest that the isolation and lack of transculturation of the Yanomami can also account for their higher bacterial diversity and prevalence.

The similarity in oral diversity levels between Yanomami and U.S. individuals, despite their clear differences in bacterial composition, is intriguing. Oral hygiene in the Yanomami is limited to occasional washing of the mouth with water. This factor combined with the ancestral habit of chewing tobacco constantly and from early ages (28) might explain why the oral microbiota is equally diverse between these populations.

In the highly diverse skin microbiota of the Yanomami, the presence of environmental bacterial taxa is consistent with their higher body environmental exposure in relation to westernized people who wear clothes and spend much time indoors. The extent to which this environmental exposure contributes to the observed increased skin diversity is currently not known.

Because the Yanomami have remained relatively isolated for more than 11,000 years since the arrival of their ancestors in the Americas (29), genomic divergence from Western strains is to be expected. Assuming a doubling rate of once per hour, an estimated 100 million bacterial generations would have occurred with little gene migration in the Yanomami microbiome, in relation to the much higher rates in the westernized world. The bacterial archive gathered from these unique specimens will be of great value to study evolution of human-associated microbial strains.

Even with no known exposure to pharmacologic-dose antibiotics, AR genes were present in the microbiome of these uncontacted Amerindians. AR genes against natural antibiotics could have entered the human microbiota through exchange with antibiotic-producing soil microbes and neighboring microbes (30, 31) or evolved in a soil-dwelling ancestor of a human commensal. Resistance to later-generation antibiotics may represent cross-resistance or a housekeeping gene that provides resistance only in a novel regulatory context (32). Alternatively, AR genes could have reached the Yanomami via traded objects or a chain of

human contact that reached antibiotic-exposed populations. These routes would require survival of commensal or cosmopolitan microbes in the environment during transfer and/or horizontal gene transfer into the human microbiota multiple times in the recent past (33). Regardless of their origin, the AR genes have been maintained in the absence of apparent antibiotic selection pressure in this very diverse Amerindian microbiome. The carriage of a pool of mobilizable resistance genes to next-generation antibiotics in antibiotic-naïve microbiota may help explain the rapidity with which clinical resistance has arisen to each new antibiotic class after its introduction (30, 34). Because antibiotics were introduced to the village after the first medical visit, future studies in this unique setting would help elucidate their effect on the human microbiome and resistome.

The high shared gene identity with commensal taxa of the source body site suggests that many of the AR genes are encoded by commensals. Continued carriage of microbes through industrialization may partially explain a common resistome in uncontacted and industrialized people. Our discovery of functional AR genes encoded by antibiotic-sensitive *E. coli* isolates confirms that culture-dependent screening severely underestimates the potential community resistome (9, 35, 36). Nevertheless, most of the identified AR genes were part of the *E. coli* core genome, in contrast with previous studies of traditional communities in the Peruvian Amazonas and in Bolivia (9, 37). In these populations, which have access to antibiotics and to urban areas, high levels of acquired resistance to antibiotics and a remarkable diversity of AR genes and plasmids have been detected in commensal *E. coli* isolates. Although AR genes are presumably silenced in Yanomami *E. coli* isolates of the current study, selection for mutant regulatory regions could readily occur with the introduction of antibiotic selective pressure, as has been observed clinically (34).

Our work emphasizes the value of deep characterization of microbiomes of people living ancestral life-styles, particularly if practices in industrialized societies might eradicate potentially beneficial microbes and their encoded functions. Characterization of the commensal resistome of diverse people will also inform the design and prudent deployment of antibiotics to minimize enrichment for preexisting resistance. Causation and functional consequences of microbiome changes need to be understood before functional restoration of the microbiome is possible to attempt to reverse the current global trends in metabolic and inflammatory diseases.

Materials and Methods

Subject and community characterization

The High Orinoco area of Amazonas state in Venezuela is a vast area (~80,000 km²) of jungle, which is the Venezuelan homeland for about 15,000 Yanomami people living in the country. Most of the villages are near rivers, but isolated villages still exist in mountainous areas, and they are described for the first time in the Annual Reports of the Venezuelan Ministry of Health. In 2008, an unmapped village was sighted by an army helicopter, and a medical mission landed there in 2009. Ó.N.-A. was the only author at the site. To protect the identity of the villagers, the name and coordinates of the village are not provided. The Yanomami from the uncontacted village were hunters and gatherers, with no agriculture or

domestication of livestock. Trade was evidenced by the presence of machetes, cans, or clothing that are commonly exchanged for arrows with other Yanomami.

The age of the 34 subjects was between 4 and 50 years, as estimated by Yanomami health workers on the medical team (table S1). The villagers were observed to eat wild bananas and seasonal fruits, plantain, palm hearts, and cassava. Hunting consists mostly of birds and small mammals, as well as small crabs, frogs, and small fish from nearby streams, and occasionally peccary, monkey, and tapir. Water is collected from a stream about 5-min walking distance from the village. They also knew the word *Medicina* from other Yanomami, but the only mission with a medical dispensary (Mahecoto) is estimated to be at least 2 weeks by foot through mountainous terrain.

For comparative purposes, we use data from our published studies of the Guahibo Amerindians and Malawians (11), from villages already in transition toward urban lifestyles, with domestication of animals, agricultural practices, some degree of market economy, medical outposts, and access to towns.

Human specimens

Volar forearm skin, oral mucosa, and fecal samples were collected before the health team vaccinated children and administered antibiotics. Samples were immediately frozen in liquid nitrogen and remained frozen until DNA extraction and bacterial culturing were performed.

Samples were cultured on diverse media under aerobic and anaerobic conditions (see below). DNA was also extracted from samples and bacterial cultures from the Yanomami and from five Puerto Rican individuals (to examine their resistome), using the PowerSoil DNA Isolation Kit (MO BIO) following the manufacturer's recommendations, with the addition of 30-min incubation after the addition of C6 buffer.

16S rRNA analysis

The V4 region of the 16S *rRNA* gene was amplified with barcoded primers and sequenced as previously described (38). α Diversity was estimated using Faith's phylogenetic diversity (39) on rarefied tables at 5000 (fecal), 1500 (oral), and 1500 (skin) sequences per sample. β Diversity was measured using unweighted UniFrac (40) on the rarefied tables. γ Diversity (total species diversity within an environment) was estimated using diversity accumulation curves for both mean bacterial genera and phylogenetic species richness as a function of the number of subjects in each population. LEfSe (14) was used with default parameters on species-level OTU tables to determine taxa that best characterize each population; only features with LDA score >2.0 were kept. For presentation purposes in Fig. 1, features with deeper taxonomic annotation (up to the genus level) are presented. A complete list of all significant features is presented in tables S7 to S9. Data from the same set of Yanomami samples generated in combination with a different group of U.S. subjects (41) using the Illumina Genome Analyzer II \times and the V2 region (27 F-338R) of the 16S *rRNA* gene were analyzed using the same parameters as described above, producing consistent results with those obtained with the V4 region (fig. S3).

Although the main driver of differences in microbiome composition among fecal samples reflected Western versus non-Western lifestyles, PCoA of the non-Western subjects alone also revealed differences between these populations, with the Yanomami appearing more similar to the Guahibo than to the Malawians (fig. S7). Because the major differences between Western and non-Western populations were due to changes within the Bacteroidetes phylum (in particular changes in the relative abundances of the *Bacteroides* and *Prevotella*), we analyzed non-Bacteroidetes OTUs alone. PCoA analysis of unweighted UniFrac distances and estimates of bacterial diversity (Faith's phylogenetic diversity) confirm that taxa outside of the Bacteroidetes phylum are also responsible for the differences between the Western and non-Western fecal microbiota (fig. S8).

Normalized prevalence/abundance plots were used to determine whether the proportion of subjects carrying an OTU is higher in a population regardless of its abundance: For each OTU, a curve was constructed measuring the percentage of a population that carries the OTU at at least a given abundance (normalized over the total abundance of the OTU).

Age correlation with bacterial diversity (Faith's phylogenetic diversity) was examined in all populations and body sites, and only diversity in Malawian fecal samples had a significant but weak positive correlation with age (fig. S9). This was expected in our village because sampled subjects were all older than 4 years, when the fecal microbiota already exhibits an adult-like structure (11).

We expanded the scope of cultures and degree of westernization by performing a meta-analysis combining our study with data from subjects living in urban (United States, Italy), agrarian (Guahibo, Malawi), and hunter-gatherer (Hazda, Tanzania) cultures (11, 42) (fig. S1). Data for Italian and Tanzania subjects were generated using the V4 region and a 454 GS-FLX Titanium sequencer, which resulted in longer reads than those obtained with Illumina technologies. Because there is only a short overlap [~ 40 base pairs (bp)] in the sequenced region from these two technologies, we did not trim sequence length for either data set, and note that this meta-analysis has some vulnerability to study effects because of a lack of fully overlapping sequence data. OTU tables were rarefied at 1000 sequences per sample, and supervised classification was performed using random forests with an out of bag error model.

Metagenomic prediction and validation

We used PICRUSt 1.0.0 (18) to predict metagenomic content from the 16S *rRNA* data and STAMP to perform pathway analysis STAMP (43). Because PICRUSt predictions rely on existing sequenced genomes that are mainly derived from Western populations, we first compared the sequenced fecal metagenomes of a Western ($n = 18$) and a non-Western ($n = 10$) population (11) against their PICRUSt-predicted metagenomes. Metagenomic reads were mapped against KEGG Orthology (KO)-annotated amino acid sequences as previously described (44), and PICRUSt-predicted metagenomes were estimated using default parameters. To determine the extent to which PICRUSt accuracy is affected by biases toward Western reference genomes, we also calculated the weighted nearest sequenced taxon index (NSTI), which measures the average branch length between OTUs in a sample and a reference genome (higher scores indicating longer distances and thus a less

characterized microbial environment). The average Spearman correlation was 0.75 (± 0.02) in the Guahibo Amerindians (fig. S2E), close to 0.78 (± 0.01) in the U.S. subjects, and to the 0.80 correlation reported in the original PICRUSt paper (18). Mean NSTI values are not significantly different ($P = 0.158$, t test), indicating that the lower correlation between the predicted and sequenced Guahibo metagenomes is not necessarily due to biases in reference genomes toward those found in Western populations.

In addition, we performed shotgun metagenomic sequencing on 8 fecal and 14 oral samples. Reads passing quality filtering were mapped as previously described (44). An average of $12.9 \pm 8.7\%$ of the reads were assigned, similar to or higher than values reported elsewhere [for example, 6.1% in (45)]. Coverage per sample was highly uneven, and we filtered out low-quality samples as follows. We constructed four subsets of the samples for which both 16S and metagenomic data were available: all samples (“all”), samples with 10,000 mapped metagenomic sequences (“10 K”), 20,000 mapped sequences (“20 K”), and 100,000 mapped sequences (“100 K”). For each of these subsets, we performed 100 rarefactions at increasing sequencing depth levels (all: 100/500/1000/5000; 10 K, 20 K, 100 K: 1000/5000/10,000) and calculated Bray-Curtis similarity and PCoA. For each sample subset and each of the 200 (100 for 16S and 100 for metagenomic) PCoA plots thus obtained, we performed Procrustes analysis (46). We found that the 10 K subset achieved the best fit at any rarefaction level (fig. S10), as shown by the lower average M^2 (goodness of fit between 16S and metagenomic data, with lower values indicating better fit), and samples with less than 10,000 reads mapped (F23, F36, F49, F50, and O18) were excluded from further analyses.

Pearson correlation between the sequenced and PICRUSt metagenome was high except for oral samples O4 and O13 (fig. S11A). In addition, we performed Procrustes analysis using Bray-Curtis distances on 100 rarefied tables for sequenced and PICRUSt-predicted metagenomes. One hundred Monte Carlo simulations were additionally performed on each of the PICRUSt versus shotgun comparisons to estimate a P value. All permutations were significant ($P < 0.01$) and M^2 values were generally low (0.27 ± 0.01), further indicating that there is a good fit between shotgun and PICRUSt data (fig. S11, B and C).

We investigated whether, despite the low and uneven coverage obtained, bacterial contigs could be assembled from the Yanomami data. We used velvet (47) to assemble the metagenomic reads into contigs ($k = 31$ to 91). The obtained assembly was fragmentary and composed mostly of short contigs (fig. S12A). Bacterial taxonomy was assigned to each contig by aligning them to IMG genomes using nucmer (48) (minimum identity $>90\%$ over at least 80% of the query length). Contigs were clustered using CONCOCT (49), and cluster quality was assessed on the basis of the taxonomic assignments for contigs within a cluster using scripts provided in CONCOCT. Although most clusters could not be unequivocally assigned to a single species, we could confirm some of the taxa identified in the Yanomami samples through 16S sequencing (table S10). Cluster quality was measured for all k -mer lengths and at three minimum contig lengths (100, 200, and 500 bp) using NMI (normalized mutual information), F_1 score, Rand index, and adjusted Rand index (fig. S12B). A subset of k -mer values were selected for each contig length upon inspection of the quality metrics, and a PCoA plot was generated for clusters obtained with those k -mer sizes and contig

lengths (fig. S12C). Overall, these results indicate that we might lack coverage to faithfully reconstruct bacterial genomes using current data. The functional uniqueness of the microbiota in these uncontacted Amerindians may be a result of their isolation for more than 11,000 years since the arrival of their ancestors in the Americas, and genomic divergence from bacterial strains found in Western populations is therefore to be expected. Assuming a doubling rate of once per hour to once per day, an estimated 4 to 100 million bacterial generations would have occurred with little gene migration in the Yanomami microbiome, in contrast to the higher rates in the westernized world.

Cultures and *E. coli* analyses

We isolated 1097 bacterial strains from frozen fecal or oral samples from the uncontacted Amerindians to provide an archive to study human microbial evolution. Specimens were cultured in seven different media [blood, chocolate, McConkey, Columbia, trypticase soy agar (TSA), CDC anaerobic blood agar (CDC), and Brucella Laked Blood Agar with Kanamycin and Vancomycin (LKV)] under aerobic and anaerobic (jars, bags, and anaerobic chamber) conditions at 37°C.

After 16S *rRNA* sequencing of 329 isolates (table S11), we characterized 24 *E. coli* isolates from fecal samples by MLST of seven housekeeping genes (19). MLST analysis was performed using the scheme at the University College of Cork (UCC) *E. coli* MLST Web site (<http://mlst.ucc.ie>), which is based on the housekeeping genes *adh*, *fumC*, *gyrB*, *icd*, *mdh*, *purA*, and *recA* (19). Polymerase chain reaction (PCR) was performed according to protocol (http://mlst.ucc.ie/mlst/dbs/Ecoli/documents/primersColi_html). All PCR products were purified using the QIAquick PCR Purification Kit (Qiagen) and sequenced in both directions. Nucleotide sequences were quality-controlled, edited, aligned, and trimmed with the SeqMan (DNASTAR Inc.). All nucleotides within the consensus sequence template were supported by at least two sequence chromatograms. A unique allele number was assigned to each distinct sequence within a locus, and a distinct sequence type (ST) number was assigned to each unique combination of alleles. Allele and ST assignments were made and deposited at the publicly accessible *E. coli* UCC MLST database (<http://mlst.ucc.ie/mlst/dbs/Ecoli>) (19). Genetic relationships of Amerindian *E. coli* isolates were assessed by comparing allelic profiles of Amerindian isolates with existing profiles present in the MLST database. The BURST clustering algorithm was used to define clonal complexes (CCs) of related isolates derived from a common ancestor, the patterns of descent linking them together, and the ancestral genotype (50, 51). Isolates were grouped into CCs if they differed at no more than one locus from at least one other member of the group (51). Founder genotypes of CCs were defined as the ST of the CC with the highest number of neighboring single-locus ST variants. All samples were confirmed to be *E. coli* by sequencing the 16S *rRNA* gene. Isolates belonged to 19 STs (table S12), of which 12 were novel, indicating that uncontacted people harbor unique strains. eBURST comparison of 19 STs found in Amerindians with that of the 3544 STs that represent the full global diversity of this species revealed that our isolates represented a broad diversity. This analysis also revealed that most of our isolates were distributed in six CCs, with three STs appearing as singletons (ST-3316, ST-164, and ST-3415) (table S12). These singletons were assigned to CC-10 (differing in less than one locus from at least one other member of the group) when less stringent conditions were used

(that is, STs sharing five or more alleles). Two CCs in particular had more than five Amerindian isolates (CC-10 and CC-58, respectively). These data suggest ongoing strain or genetic transfer among the *E. coli* populations of the individuals in our study, probably facilitated by the fact that all lived in the same village with limited external contact.

The ECOR collection of 72 reference strains (52), which represent the breadth of genetic diversity within the species *E. coli*, was included for phylogenetic analysis. In-frame–concatenated sequences were aligned using MUSCLE (53), and the ML tree was reconstructed with the HKY85 model of DNA evolution (100 bootstrap replicates) using PhyML (54). In PhyML, the transition/transversion ratio, proportion of invariant sites, and the γ shape parameter were estimated, and subtree-pruning and regrafting and nearest-neighbor interchange tree topology search methods were used. *Escherichia fergusonii* was used as an outgroup to root the *E. coli* phylogeny. Gene sequences from all seven MLST loci were also analyzed using ClonalFrame version 1.2 (55), a Bayesian method of constructing evolutionary histories that takes both mutation and recombination into account. Five independent runs were performed, each consisting of 100,000 iterations, including a 50,000 burn-in. Analyses from five independent runs were concatenated to construct a 50% majority-rule consensus tree. MEGA5 (56) was used to draw the consensus phylogenetic tree.

AR phenotypes in *E. coli* from Amerindians

Of the 329 fecal bacterial isolates (table S11), 131 were *E. coli*. Of these, all were tested for susceptibility to 23 antibiotics.

MDA of metagenomic DNA

We performed MDA on fecal and oral metagenomic DNA from Amerindian subjects 3, 5, 6, and 23 and the five Puerto Rican subjects with the illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare Life Sciences). The recommended protocol was followed with a 2-hour incubation. We pooled three to six reactions per sample to reduce MDA template amplification bias.

Creation of metagenomic libraries

Small-insert metagenomic libraries were constructed from the MDA-amplified metagenomic DNA in the vector pZE21 as previously described (table S5) (20, 31, 57). The DNA fragments cloned into pZE21 contain several bacterial genes and are situated downstream of the constitutively active p_{LTetO-1} promoter, enabling expression and phenotypic selection of the shotgun-cloned fragments.

Creation of *E. coli* isolate genomic libraries

The *E. coli* isolate genomic libraries were created from 38 isolates from the 11 individuals who provided fecal samples. The number of isolates per individual ranged from 1 to 7. Genomic DNA (500 ng each) from 38 *E. coli* isolates was combined into two pools, each containing DNA from 19 isolates. Genomic DNA was not amplified with MDA. Small-insert libraries (libraries A and B) were created from the pooled genomic DNA using the procedure above (table S2).

Functional selection of libraries for AR

Metagenomic and genomic libraries were plated on Mueller-Hinton agar supplemented with kanamycin (50 µg/ml) (the selection marker in the pZE21 vector) and another antibiotic (table S6). Puerto Rican and *E. coli* isolate libraries were not selected against nitrofurantoin. Antibiotic concentrations were clinically relevant concentrations recommended by the Clinical and Laboratory Standards Institute or the minimum inhibitory concentration of the library host MegaX transformed with self-ligated pZE21. In all cases, antibiotic concentrations inhibited MegaX transformed with self-ligated pZE21. Surviving colonies were pooled. In some cases, resistant colonies were confirmed by replating on solid media (F6 cefepime and O23 tetracycline selections) or by inoculating into liquid media (F6 cefepime, 238 penicillin, F3, F6, F23, T1001 tetracycline, O3 chloramphenicol, 238, 337, T1001 piperacillin-tazobactam, O23, T1001, library B ceftazidime, and library A and B aztreonam selections) containing the selective antibiotic at the same concentration. Reselected clones were used as the template for sequencing.

Sequencing and assembly of resistance-conferring inserts

Resistance-conferring inserts were isolated from pooled surviving colonies via PCR with vector-specific primers, barcoded, and sequenced in parallel with the Illumina HiSeq 2000 (2 × 101-bp reads). Reads were assembled into contigs, and open reading frames were identified with PARFuMS and then annotated with Resfams (31, 58). The sequences of the F6 cefepime, O23 tetracycline, and *E. coli* library A ceftazidime-resistant inserts were determined by paired-end Sanger sequencing performed at the Protein and Nucleic Acid Chemistry Laboratory (PNAACL) at Washington University School of Medicine. For F6_AZ_9, F6_CT_1, O3_CZ_2_1, O3_CZ_2_2, O3_TE_1, O23_TE_1, F23_TE_1, F6_CH_4, O23_CH_19, and F23_CH_11, Sanger sequencing was performed at PNAACL or at Genewiz with specific primers to confirm portions of the sequence, and the Sanger sequence was used for all analyses.

Computational filtering of assembled contigs from metagenomic libraries

Nonspecific amplification is frequently observed for MDA with commercial MDA reagents, particularly with small amounts of starting template (59). A total of 171 ng of DNA was amplified from three no-template control (NTC) reactions performed in parallel with the Amerindian sample amplifications, ~15× less output DNA per reaction than for the fecal and oral samples (57 ± 25 ng versus 854 ± 67 ng). There was no detectable amplification in the NTCs for the Puerto Rico sample amplifications (<0.25 ng each). We shotgun-sequenced 84 ng of the Amerindian NTC product without antibiotic selection and assembled the reads with the same procedure as the sample reads. To ensure the authenticity of the genes identified in the metagenomic functional selections, we excluded any sample contig from analysis that aligned to an NTC contig with at least 99% nucleotide identity over at least 200 bp (113 Amerindian and 4 Puerto Rican contigs). Additionally, we excluded any sample contig whose best hit in NCBI nt (16 August 2013) was a vector or mouse/human DNA, provided that the alignment was at least 200 bp long and had 99% sequence identity (four and two Amerindian contigs, respectively, and no Puerto Rican contigs). *E. coli* isolate

libraries were not filtered because they did not undergo MDA. We did not consider any contigs under 500 bp.

AR gene identification

From the contigs that passed computational filtering, AR genes were identified on the basis of Resfams annotations specific to that antibiotic class, similarity to known resistance genes (for transporters), or subcloning and functional confirmation. The genes F23_TE_1:69-1403, O23_TE_1:1-1845, O23_CH_19:550-804, and F23_CH_11:1298-1636 were PCR-amplified from their respective antibiotic selections with primers encompassing the entire gene. They were ligated into pZE21 at an optimal distance from the RBS (ribosome binding site) and transformed into *E. coli* DH10B. The resistance phenotype was confirmed by plating on the selective antibiotic at the same concentration used in the initial selection and amplification of the fragment from surviving colonies with specific primers.

Alignment to NCBI

AR genes were compared to NCBI nr (non-redundant protein database) (downloaded 15 September 2014; including uncultured/environmental sample sequences) via blastx. The top hit was the hit with the lowest *e* value; any hits with the same *e* value and score were also considered top hits. Alignments 200 bp or shorter were excluded from analysis. Global %ID (table S4) was calculated from estwise (Wise2 package) sequence alignments. To compare the context of AR genes in contigs isolated from Amerindian metagenomes and in industrialized settings, genes and the surrounding contigs were compared to NCBI nt (15 August 2013).

Tree building for Amerindian PBPs

For each gene, the top blastx hit in NCBI nr and any hit with amino acid identity 99% and coverage 90% of the gene was downloaded from NCBI. Proteins were clustered with cd-hit at 99.5% identity (60), and the representative sequence for each cluster was used for tree building. Functionally isolated AR genes were translated with EMBOSS Transeq (61, 62). The phylogenetic tree was created with the Phylogeny.fr pipeline (63). Proteins were aligned with MUSCLE (Full Mode) (53), and PhyML3.0 (54, 64) was used to create the phylogenetic tree (default parameters, default substitution model, SH-like approximate likelihood ratio test for branch support), which was visualized with Geneious version 7.0 (Biomatters).

Alignment of functionally selected AR genes to the HMP and MetaHIT

AR genes were compared to the HMP (65) and MetaHIT (66) shotgun assemblies (HMP: 749 assemblies, <http://www.hmpdacc.org/HMASM/>; MetaHIT: 124 assemblies, NCBI/European Bioinformatics Institute accession ERA000116, http://www.bork.embl.de/~arumugam/Qin_et_al_2010/) via blastn. The reported sequence identity for each gene is that of the top alignment (lowest *e* value) in that database. Alignments 200 bp or shorter were excluded from analysis.

Alignment to *E. coli* MG1655

All AR genes isolated from the *E. coli* genomic libraries were compared to *E. coli* str. K-12 substr. MG1655 (RefSeq: NC_000913.3) on 19 September 2014 to determine the gene identity. All genes with MG1655 alignments (39 of 44 genes) aligned with 96% nucleotide identity.

Filtering of human sequences from shotgun reads

Human sequences were filtered from the forward-only trimmed shotgun reads by mapping to the human reference genome (GRCh38) with DeconSeq (67). Filtered reads were uploaded to the NCBI Sequence Read Archive (SRA) (SRP049631) and mapped to resistance genes.

Comparison of shotgun reads to functionally selected resistance genes

All functionally selected AR genes from the Amerindian and Puerto Rican oral and fecal metagenomic libraries and the Amerindian *E. coli* isolate genomic DNA libraries were clustered with cd-hit-est at 100% nucleotide identity over the length of the smaller gene (60). Forward-only trimmed and filtered shotgun reads from Amerindian fecal and oral microbiota were mapped to the clustered AR genes with bowtie2 (end-to-end alignment, one alignment per read) (68). Alignments were filtered to those with an alignment score $-0.3278 \times$ the length of the read ($\sim 95\%$ ID). Gene abundance in the metagenome was calculated as the number of reads mapping to the gene by the length of the gene in kilobases by millions of reads mapped. A Bray-Curtis dissimilarity matrix was calculated in R (package vegan, function vegdist) (69) from the gene \times sample abundance table. Samples were clustered on the dissimilarity matrix with UPGMA (function hclust). Heatmaps of gene abundance were generated with R packages NMF and RColorBrewer (70, 71). Statistical differences in abundance between categories were calculated with Wilcoxon rank sum tests and corrected for multiple hypotheses with a Bonferroni correction.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the Venezuelan “Fuerza Armada Nacional Bolivariana” and the members of the Health expedition, including medical technologists C. Ordaz, whose support made possible the collecting of these unique specimens. We are also grateful to the people of the newly contacted village for their trust and our shared desire that the unavoidable contact with our culture will bring health benefits and protection to their people. We thank M. Charbonneau, H. J. Lee, and J. Gu for optimization of the MDA metagenome amplification protocols. D. Wendel, J. Stombaugh, and G. Ackermann assisted with data processing. D. Berg-Lyons, G. Humphrey, and C. Lauber assisted with sample processing.

Funding: This work was supported in part by the C&D Fund (M.G.D.-B.), the Emch Fund (M.G.D.-B.), NIH SCORE-PES RISE (M.G.D.-B.), SUCCESS (J.C.C.), Helmsley Charitable Trust HSP-007 (J.C.C.), NIH National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) DK062429 (J.C.C.), NAKFI Synthetic Biology—SB2 (G.D. and R.K.), NIH Director's New Innovator Award NIDDK DP2-DK098089 (G.D.), NIH National Institute of General Medical Sciences R01-GM099538 (G.D.), Washington University I-CARES award (G.D.), UH2AR057506 and the Diane Belfer Program for Human Microbial Ecology (M.J.B.), National Defense Science and Engineering Graduate fellowship (E.C.P.), and Howard Hughes Medical Institute Early Career Scientist (R.K.).

The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

References

1. Henao-Mejia J, Elinav E, Jin C, Hao L, Mehal WZ, Strowig T, Thaiss CA, Kau AL, Eisenbarth SC, Jurczak MJ, Camporez JP, Shulman GI, Gordon JI, Hoffman HM, Flavell RA. Inflammation-mediated dysbiosis regulates progression of NAFLD and obesity. *Nature*. 2012; 482:179–185. [PubMed: 22297845]
2. Cho I, Yamanishi S, Cox L, Methé BA, Zavadil J, Li K, Gao Z, Mahana D, Raju K, Teitler I, Li H, Alekseyenko AV, Blaser MJ. Antibiotics in early life alter the murine colonic microbiome and adiposity. *Nature*. 2012; 488:621–626. [PubMed: 22914093]
3. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R, Gordon JI. A core gut microbiome in obese and lean twins. *Nature*. 2009; 457:480–484. [PubMed: 19043404]
4. Diaz HR, Wang S, Anuar F, Qian Y, Björkholm B, Samuelsson A, Hibberd ML, Forssberg H, Pettersson S. Normal gut microbiota modulates brain development and behavior. *Proc Natl Acad Sci U S A*. 2011; 108:3047–3052. [PubMed: 21282636]
5. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*. 2006; 444:1027–1031. [PubMed: 17183312]
6. Koren O, Goodrich JK, Cullender TC, Spor A, Laitinen K, Bäckhed HK, Gonzalez A, Werner JJ, Angenent LT, Knight R, Bäckhed F, Isolauri E, Salminen S, Ley RE. Host remodeling of the gut microbiome and metabolic changes during pregnancy. *Cell*. 2012; 150:470–480. [PubMed: 22863002]
7. Ministerio del Poder Popular para la Salud. Memoria 2008–2011. Caracas, Venezuela: 2009–2012.
8. Sommer MO, Dantas G, Church GM. Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science*. 2009; 325:1128–1131. [PubMed: 19713526]
9. Bartoloni A, Pallecchi L, Rodríguez H, Fernandez C, Mantela A, Bartalesi F, Strohmeyer M, Kristiansson C, Gotuzzo E, Paradisi F, Rossolini GM. Antibiotic resistance in a very remote Amazonas community. *Int J Antimicrob Agents*. 2009; 33:125–129. [PubMed: 18947984]
10. Warinner C, Rodrigues JF, Vyas R, Trachsel C, Shved N, Grossmann J, Radini A, Hancock Y, Tito RY, Fiddyment S, Speller C, Hendy J, Charlton S, Luder HU, Salazar-García DC, Eppler E, Seiler R, Hansen LH, Castruita JA, Barkow-Oesterreicher S, Teoh KY, Kelstrup CD, Olsen JV, Nanni P, Kawai T, Willerslev E, von Mering C, Lewis CM Jr, Collins MJ, Gilbert MT, Rühli F, Cappellini E. Pathogens and host immunity in the ancient human oral cavity. *Nat Genet*. 2014; 46:336–344. [PubMed: 24562188]
11. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP, Heath AC, Warner B, Reeder J, Kuczynski J, Caporaso JG, Lozupone CA, Lauber C, Clemente JC, Knights D, Knight R, Gordon JI. Human gut microbiome viewed across age and geography. *Nature*. 2012; 486:222–227. [PubMed: 22699611]
12. American Gut. <http://github.com/biocore/American-Gut>
13. Pontzer H, Raichlen DA, Wood BM, Mabulla AZ, Racette SB, Marlowe FW. Huntergatherer energetics and human obesity. *PLOS One*. 2012; 7:e40503. [PubMed: 22848382]
14. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C. Metagenomic biomarker discovery and explanation. *Genome Biol*. 2011; 12:R60. [PubMed: 21702898]
15. Groth I, Schumann P, Schutze B, Augsten K, Stackebrandt E. *Knoellia sinensis* gen. nov., sp. nov. and *Knoellia subterranea* sp. nov., two novel actinobacteria isolated from a cave. *Int J Syst Evol Microbiol*. 2002; 52:77–84. [PubMed: 11837319]
16. Shin NR, Roh SW, Kim MS, Jung MJ, Whon TW, Bae JW. *Knoellia locipacati* sp. nov., from soil of the Demilitarized Zone in South Korea. *Int J Syst Evol Microbiol*. 2012; 62:342–346. [PubMed: 21421930]
17. Ward NL, Challacombe JF, Janssen PH, Henrissat B, Coutinho PM, Wu M, Xie G, Haft DH, Sait M, Badger J, Barabote RD, Bradley B, Brettin TS, Brinkac LM, Bruce D, Creasy T, Daugherty

- SC, Davidsen TM, DeBoy RT, Detter JC, Dodson RJ, Durkin AS, Ganapathy A, Gwinn-Giglio M, Han CS, Khouri H, Kiss H, Kothari SP, Madupu R, Nelson KE, Nelson WC, Paulsen I, Penn K, Ren Q, Rosovitz MJ, Selengut JD, Shrivastava S, Sullivan SA, Tapia R, Thompson LS, Watkins KL, Yang Q, Yu C, Zafar N, Zhou L, Kuske CR. Three genomes from the phylum *Acidobacteria* provide insight into the lifestyles of these microorganisms in soils. *Appl Environ Microbiol.* 2009; 75:2046–2056. [PubMed: 19201974]
18. Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepille DE, Vega TRL, Knight R, Beiko RG, Huttenhower C. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol.* 2013; 31:814–821. [PubMed: 23975157]
 19. Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, Karch H, Reeves PR, Maiden MC, Ochman H, Achtman M. Sex and virulence in *Escherichia coli*: An evolutionary perspective. *Mol Microbiol.* 2006; 60:1136–1151. [PubMed: 16689791]
 20. Forsberg KJ, Patel S, Gibson MK, Lauber CL, Knight R, Fierer N, Dantas G. Bacterial phylogeny structures soil resistomes across habitats. *Nature.* 2014; 509:612–616. [PubMed: 24847883]
 21. Contreras M, Costello EK, Hidalgo G, Magris M, Knight R, Dominguez-Bello MG. The bacterial microbiota in the oral mucosa of rural Amerindians. *Microbiology.* 2010; 156:3282–3287. [PubMed: 20847007]
 22. Lester CH, Frimodt-Møller N, Sørensen TL, Monnet DL, Hammerum AM. In vivo transfer of the *vanA* resistance gene from an *Enterococcus faecium* isolate of animal origin to an *E. faecium* isolate of human origin in the intestines of human volunteers. *Antimicrob Agents Chemother.* 2006; 50:596–599. [PubMed: 16436715]
 23. Cox LM, Yamanishi S, Sohn J, Alekseyenko AV, Leung JM, Cho I, Kim SG, Li H, Gao Z, Mahana D, Zárate Rodríguez JG, Rogers AB, Robine N, Loke P, Blaser MJ. Altering the intestinal microbiota during a critical developmental window has lasting metabolic consequences. *Cell.* 2014; 158:705–721. [PubMed: 25126780]
 24. Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N, Knight R. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci USA.* 2010; 107:11971–11975. [PubMed: 20566857]
 25. Chesson P. Mechanisms of maintenance of species diversity. *Annu Rev Ecol Syst.* 2000; 31:343–366.
 26. Leibold MA, Holyoak M, Mouquet N, Amarasekare P, Chase JM, Hoopes MF, Holt RD, Shurin JB, Law R, Tilman D, Loreau M, Gonzalez A. The metacommunity concept: A framework for multi-scale community ecology. *Ecol Lett.* 2004; 7:601–613.
 27. Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Seedorf H, Goodman AL, Clemente JC, Knight R, Heath AC, Leibel RL, Rosenbaum M, Gordon JI. The long-term stability of the human gut microbiota. *Science.* 2013; 341:1237439. [PubMed: 23828941]
 28. Charlton A. Medicinal uses of tobacco in history. *J R Soc Med.* 2004; 97:292–296. [PubMed: 15173337]
 29. Salzano, FM.; Callegari-Jacques, SM. South American Indians: A Case Study in Evolution. Clarendon Press; Oxford: 1988.
 30. D'Costa VM, King CE, Kalan L, Morar M, Sung WW, Schwarz C, Froese D, Zazula G, Calmels F, Debruyne R, Golding GB, Poinar HN, Wright GD. Antibiotic resistance is ancient. *Nature.* 2011; 477:457–461. [PubMed: 21881561]
 31. Forsberg KJ, Reyes A, Wang B, Selleck EM, Sommer MO, Dantas G. The shared antibiotic resistome of soil bacteria and human pathogens. *Science.* 2012; 337:1107–1111. [PubMed: 22936781]
 32. Dantas G, Sommer MO. Context matters—The complex interplay between resistome genotypes and resistance phenotypes. *Curr Opin Microbiol.* 2012; 15:577–582. [PubMed: 22954750]
 33. Martínez JL, Coque TM, Baquero F. What is a resistance gene? Ranking risk in resistomes. *Nat Rev Microbiol.* 2015; 13:116–123. [PubMed: 25534811]
 34. Tracz DM, Boyd DA, Hizon R, Bryce E, McGeer A, Ofner-Agostini M, Simor AE, Paton S. MR; Canadian Nosocomial Infection Surveillance Program Mulvey. *ampC* gene expression in promoter

- mutants of cefoxitin-resistant *Escherichia coli* clinical isolates. FEMS Microbiol Lett. 2007; 270:265–271. [PubMed: 17326753]
35. Houndt T, Ochman H. Long-term shifts in patterns of antibiotic resistance in enteric bacteria. Appl Environ Microbiol. 2000; 66:5406–5409. [PubMed: 11097921]
 36. Walson JL, Marshall B, Pokhrel BM, Kafle KK, Levy SB. Carriage of antibiotic-resistant fecal bacteria in Nepal reflects proximity to Kathmandu. J Infect Dis. 2001; 184:1163–1169. [PubMed: 11598839]
 37. Bartoloni A, Bartalesi F, Mantela A, Dell'Amico E, Roselli M, Strohmeier M, Barahona HG, Barrón VP, Paradisi F, Rossolini GM. High prevalence of acquired antimicrobial resistance unrelated to heavy antimicrobial consumption. J Infect Dis. 2004; 189:1291–1294. [PubMed: 15031799]
 38. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G, Knight R. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 2012; 6:1621–1624. [PubMed: 22402401]
 39. Faith DP. Conservation evaluation and phylogenetic diversity. Biol Conserv. 1992; 61:1–10.
 40. Lozupone C, Knight R. UniFrac: A new phylogenetic method for comparing microbial communities. Appl Environ Microbiol. 2005; 71:8228–8235. [PubMed: 16332807]
 41. Song SJ, Lauber C, Costello EK, Lozupone CA, Humphrey G, Berg-Lyons D, Caporaso JG, Knights D, Clemente JC, Nakielnny S, Gordon JI, Fierer N, Knight R. Cohabiting family members share microbiota with one another and with their dogs. eLife. 2013; 2:e00458. [PubMed: 23599893]
 42. Schnorr SL, Candela M, Rampelli S, Centanni M, Consolandi C, Basaglia G, Turrone S, Biagi E, Peano C, Severgnini M, Fiori J, Gotti R, De Bellis G, Luiselli D, Brigidi P, Mabula A, Marlowe F, Henry AG, Crittenden AN. Gut microbiome of the Hadza hunter-gatherers. Nat Commun. 2014; 5:3654. [PubMed: 24736369]
 43. Parks DH, Tyson GW, Hugenholtz P, Beiko RG. STAMP: Statistical analysis of taxonomic and functional profiles. Bioinformatics. 2014; 30:3123–3124. [PubMed: 25061070]
 44. Fierer N, Ladau J, Clemente JC, Leff JW, Owens SM, Pollard KS, Knight R, Gilbert JA, McCulley RL. Reconstructing the microbial diversity and function of pre-agricultural tallgrass prairie soils in the United States. Science. 2013; 342:621–624. [PubMed: 24179225]
 45. Roggenbuck M, Bærholm SI, Blom N, Bælum J, Bertelsen MF, Pontén TS, Sørensen SJ, Gilbert MT, Graves GR, Hansen LH. The microbiome of New World vultures. Nat Commun. 2014; 5:5498. [PubMed: 25423494]
 46. Muegge BD, Kuczynski J, Knights D, Clemente JC, González A, Fontana L, Henrissat B, Knight R, Gordon JI. Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. Science. 2011; 332:970–974. [PubMed: 21596990]
 47. Zerbino DR, Birney E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008; 18:821–829. [PubMed: 18349386]
 48. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. Versatile and open software for comparing large genomes. Genome Biol. 2004; 5:R12. [PubMed: 14759262]
 49. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C. Binning metagenomic contigs by coverage and composition. Nat Methods. 2014; 11:1144–1146. [PubMed: 25218180]
 50. Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG. eBURST: Inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol. 2004; 186:1518–1530. [PubMed: 14973027]
 51. Jolley KA, Feil EJ, Chan MS, Maiden MC. Sequence type analysis and recombinational tests (START). Bioinformatics. 2001; 17:1230–1231. [PubMed: 11751234]
 52. Ochman H, Selander RK. Standard reference strains of *Escherichia coli* from natural populations. J Bacteriol. 1984; 157:690–693. [PubMed: 6363394]
 53. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004; 32:1792–1797. [PubMed: 15034147]

54. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol.* 2010; 59:307–321. [PubMed: 20525638]
55. Didelot X, Falush D. Inference of bacterial microevolution using multilocus sequence data. *Genetics.* 2007; 175:1251–1266. [PubMed: 17151252]
56. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 2011; 28:2731–2739. [PubMed: 21546353]
57. Moore AM, Patel S, Forsberg KJ, Wang B, Bentley G, Razia Y, Qin X, Tarr PI, Dantas G. Pediatric fecal microbiota harbor diverse and novel antibiotic resistance genes. *PLOS One.* 2013; 8:e78822. [PubMed: 24236055]
58. Gibson MK, Forsberg KJ, Dantas G. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *ISME J.* 2015; 9:207–216. [PubMed: 25003965]
59. Blainey PC, Quake SR. Digital MDA for enumeration of total nucleic acid contamination. *Nucleic Acids Res.* 2011; 39:e19. [PubMed: 21071419]
60. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: Accelerated for clustering the next generation sequencing data. *Bioinformatics.* 2012; 28:3150–3152. [PubMed: 23060610]
61. Rice P, Longden I, Bleasby A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* 2000; 16:276–277. [PubMed: 10827456]
62. Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.* 2010; 38:W695–W699. [PubMed: 20439314]
63. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O. Phylogeny. fr: Robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* 2008; 36:W465–W469. [PubMed: 18424797]
64. Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol.* 2006; 55:539–552. [PubMed: 16785212]
65. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature.* 2012; 486:207–214. [PubMed: 22699609]
66. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto JM, Hansen T, Le Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, Sicheritz-Ponten T, Turner K, Zhu H, Yu C, Li S, Jian M, Zhou Y, Li Y, Zhang X, Li S, Qin N, Yang H, Wang J, Brunak S, Doré J, Guarner F, Kristiansen K, Pedersen O, Parkhill J, Weissenbach J, Bork P, Ehrlich SD, Wang J. MetaHIT Consortium. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature.* 2010; 464:59–65. [PubMed: 20203603]
67. Schmieder R, Edwards R. Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLOS One.* 2011; 6:e17288. [PubMed: 21408061]
68. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012; 9:357–359. [PubMed: 22388286]
69. vegan: Community Ecology Package. R package version 2.0-10. 2013
70. RColorBrewer: ColorBrewer palettes. R package version 1.0-5. 2011
71. Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics.* 2010; 11:367. [PubMed: 20598126]
72. Rideout JR, He Y, Navas-Molina JA, Walters WA, Ursell LK, Gibbons SM, Chase J, McDonald D, Gonzalez A, Robbins-Pianka A, Clemente JC, Gilbert JA, Huse SM, Zhou HW, Knight R, Caporaso JG. Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and scales to billions of sequences. *PeerJ.* 2014; 2:e545. [PubMed: 25177538]
73. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* 2012; 6:610–618. [PubMed: 22134646]
74. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014; 30:2114–2120. [PubMed: 24695404]

75. Kazimierczak KA, Flint HJ, Scott KP. Comparative analysis of sequences flanking *tet(W)* resistance genes in multiple species of gut bacteria. *Antimicrob Agents Chemother.* 2006; 50:2632–2639. [PubMed: 16870752]
76. Beaber JW, Hochhut B, Waldor MK. SOS response promotes horizontal dissemination of antibiotic resistance genes. *Nature.* 2004; 427:72–74. [PubMed: 14688795]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

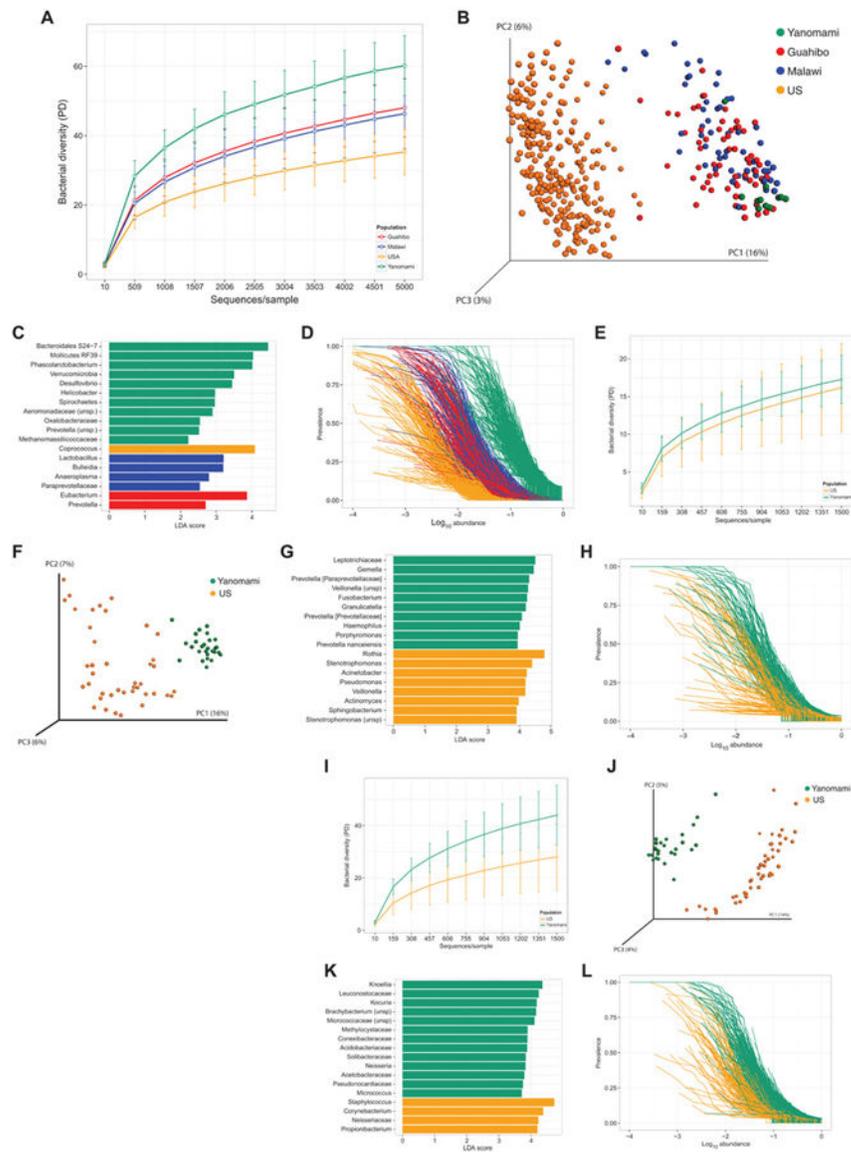


Fig. 1. Microbiota diversity in fecal, oral, and skin samples from uncontacted Yanomami in relation to other human groups

(A) Faith's phylogenetic diversity (PD) (average \pm SD) of fecal samples from Yanomami and Guahibo Amerindians, Malawians, and U.S. subjects. OTU tables rarefied at 5000 sequences per sample. Intersubpopulation differences were significant ($P < 0.001$, ANOVA with Tukey's HSD) for all but Guahibo-Malawi comparison ($P = 0.73$). (B) PCoA plot based on UniFrac distances calculated on the OTU table of fecal samples rarefied at 5000 sequences per sample. (C) Top discriminative bacteria among populations in fecal samples as determined by linear discriminant analysis (LDA) effect size (LEfSe) analysis. (D) Normalized prevalence/abundance curves for all OTUs found at 1% abundance or more in fecal samples. (E) Faith's phylogenetic diversity (average \pm SD) of oral samples from Yanomami and U.S. subjects. OTU tables rarefied at 1500 sequences per sample. Intersubpopulation differences were not significant ($P = 0.296$, ANOVA with Tukey's HSD). (F) PCoA plot based on UniFrac distances calculated on OTU tables of oral samples rarefied

at 1500 sequences per sample. **(G)** Top discriminative bacteria among populations in oral samples as determined by LEfSe analysis. **(H)** Normalized prevalence/abundance curves for all OTUs found at 1% abundance or more in oral samples. **(I)** Faith's phylogenetic diversity (average \pm SD) of skin samples from Yanomami and U.S. subjects. OTU tables rarefied at 1500 sequences per sample. Interpopulation differences were significant ($P < 0.001$, ANOVA with Tukey's HSD). **(J)** PCoA plot based on UniFrac distances calculated on OTU tables of skin samples rarefied at 1500 sequences per sample. **(K)** Top discriminative bacteria among populations in skin samples as determined by LEfSe analysis. **(L)** Normalized prevalence/abundance curves for all OTUs found at 1% abundance or more in skin samples.

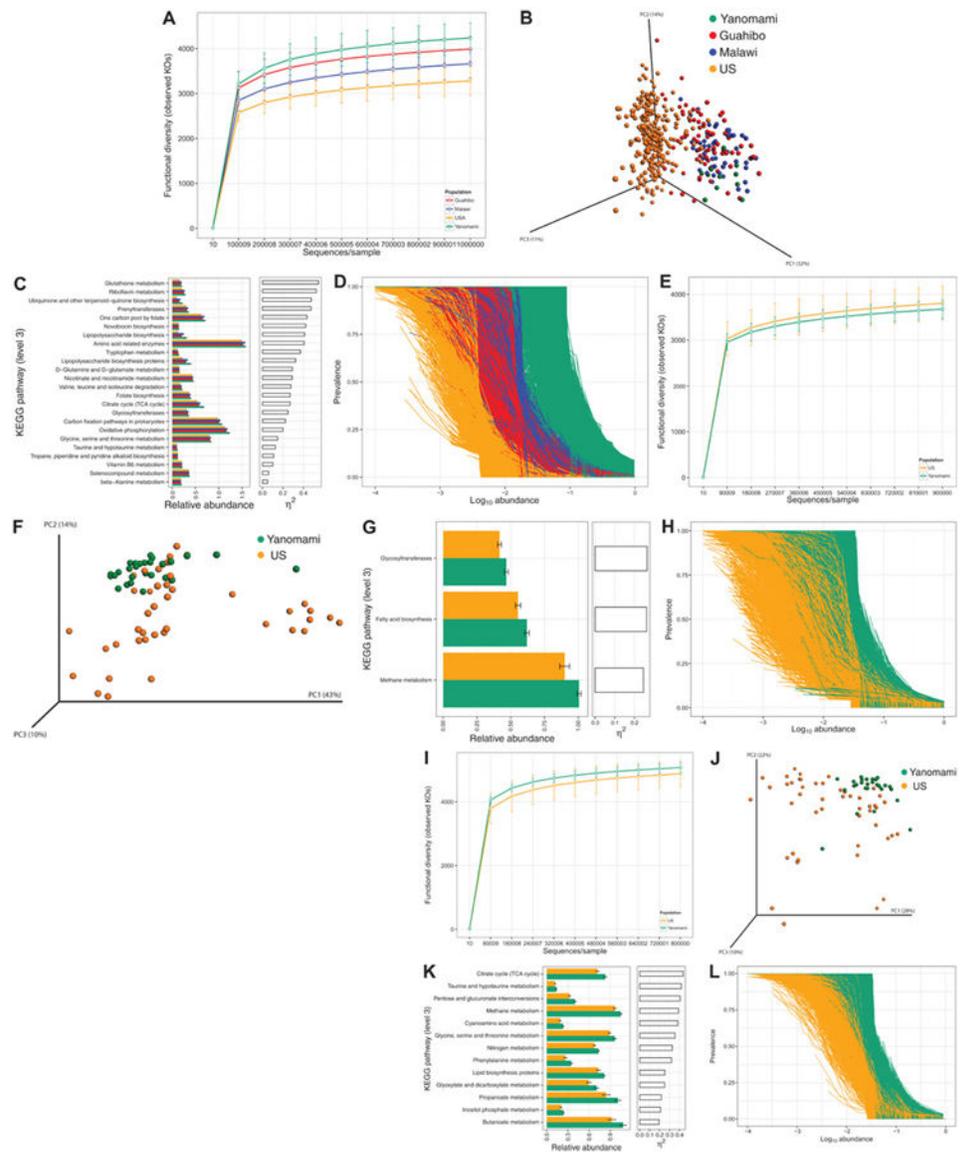


Fig. 2. Metagenomic diversity of fecal and oral samples from uncontacted Yanomami in relation to other human groups

(A) Functional diversity (average \pm SD) measured by the observed number of KEGG orthologs in fecal samples of Yanomami and Guahibo Amerindians, Malawians, and U.S. subjects. Metagenomic tables rarefied at 1 million sequences per sample. Interpopulation differences were significant for all comparisons ($P < 0.001$, ANOVA and Tukey's HSD). (B) PCoA plot based on Bray-Curtis distances calculated on the metagenomic table of fecal samples rarefied at 1 million sequences per sample. (C) Top discriminative metabolic pathways in Yanomami fecal samples as determined by STAMP; pathways ranked by effect size (η^2). (D) Normalized prevalence/abundance curves for all KOs found at 1% abundance or more in fecal samples. (E) Functional diversity (average \pm SD) measured by the observed number of KEGG orthologs in oral samples of Yanomami and Guahibo Amerindians, Malawians, and U.S. subjects. Metagenomic tables rarefied at 900,000 sequences per sample. Interpopulation differences were not significant ($P = 0.07$, t test). (F) PCoA plot

based on Bray-Curtis distances calculated on the metagenomic table of oral samples rarefied at 900,000 sequences per sample. **(G)** Top discriminative metabolic pathways in Yanomami oral samples as determined by STAMP; pathways ranked by effect size (η^2). **(H)** Normalized prevalence/abundance curves for all KOs found at 1% abundance or more in oral samples. **(I)** Functional diversity (average \pm SD) measured by the observed number of KEGG orthologs in skin samples of Yanomami and U.S. subjects. Metagenomic tables rarefied at 1 million sequences per sample. Interpopulation differences were significant for all comparisons ($P < 0.001$, ANOVA and Tukey's HSD). **(J)** PCoA plot based on Bray-Curtis distances calculated on the metagenomic table of skin samples rarefied at 1 million sequences per sample. **(K)** Top discriminative metabolic pathways in Yanomami skin samples as determined by STAMP; pathways ranked by effect size (η^2). **(L)** Normalized prevalence/abundance curves for all KOs found at 1% abundance or more in skin samples.

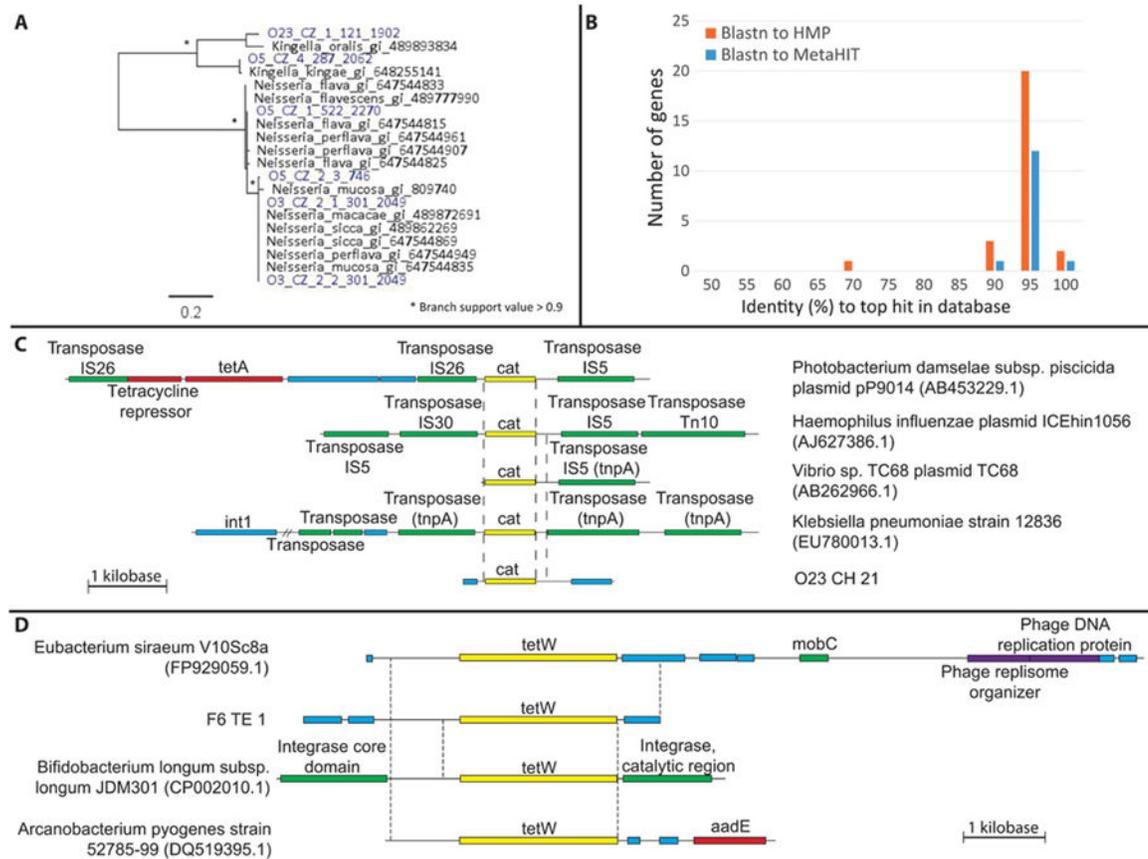


Fig. 3. Representative AR genes captured through functional metagenomic selection of Amerindian oral and fecal microbiota

(A) Maximum likelihood (ML) tree of ceftazidime-resistant PBPs from Amerindian oral microbiota (blue) and their top blastx hits in NCBI nr (black). (B) Identity (%) of the 28 AR genes to their top hits in HMP and MetaHIT (blastn). Fourteen aligned to MetaHIT, all isolated from fecal microbiota. (C) Chloramphenicol acetyltransferase (*cat*) from O23_CH_21 and homologs in NCBI nt (nucleotide database). Dashed lines indicate 99% identical sequences. Dashes indicate omitted sequence (8487..25716). (D) *tetW* from F6_TE_1 and homologs in NCBI nt. Dashed lines indicate 98 to 99% identical sequences. Sequences are to scale. Yellow, AR gene; red, other resistance genes; green, mobile genetic elements; purple, bacteriophage genes; blue, other.

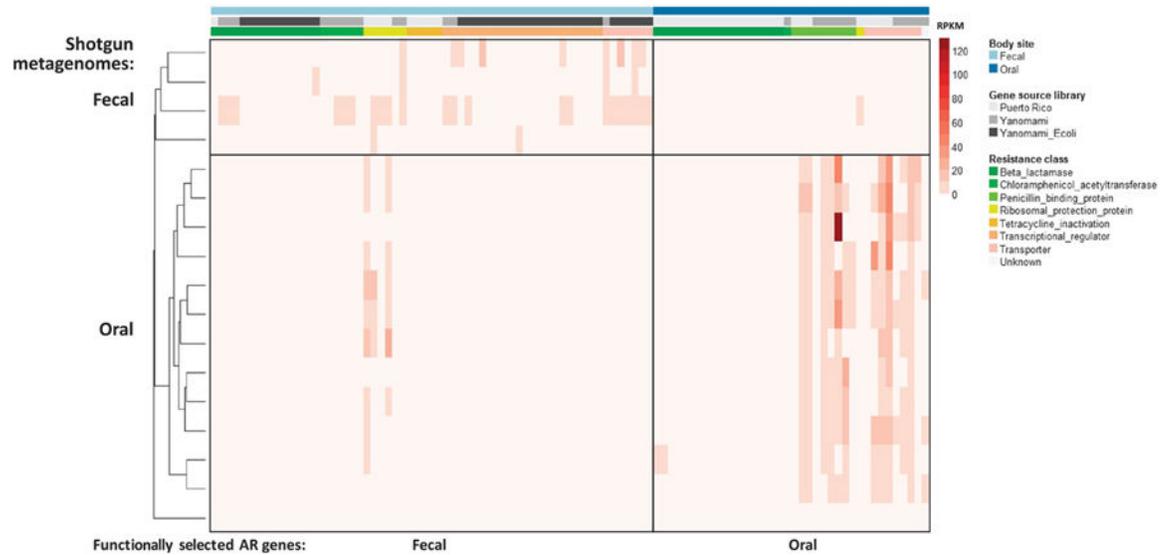


Fig. 4. Abundance of functionally selected AR genes in Amerindian fecal and oral metagenomes
 The heatmap displays the abundances in RPKM (reads per 1-kb gene sequence per million reads mapped) of functionally selected AR genes in the fecal and oral metagenomes of Amerindians. The AR genes were functionally selected from the oral and fecal microbiota of the Yanomami individuals and matched Puerto Rican controls. The shotgun-sequenced metagenomes are from the Yanomami villagers, including individuals not investigated with functional metagenomic selections. Shotgun-sequenced metagenomes were clustered by UPGMA (unweighted pair group method with arithmetic mean) on the AR gene abundance profiles. Fecal and oral microbiota cluster apart.