

ORIGINAL ARTICLE

Distinct cutaneous bacterial assemblages in a sampling of South American Amerindians and US residents

Martin J Blaser¹, Maria G Dominguez-Bello², Monica Contreras³, Magda Magris⁴, Glida Hidalgo⁴, Isidoro Estrada^{5,6}, Zhan Gao¹, Jose C Clemente⁷, Elizabeth K Costello⁸ and Rob Knight^{7,9}

¹Departments of Medicine and Microbiology, New York University Langone Medical Center, New York, NY, USA; ²Department of Biology, University of Puerto Rico, San Juan, PR, USA; ³Instituto Venezolano de Investigaciones Cientificas (IVIC), Caracas, Venezuela; ⁴SACAICET, Puerto Ayacucho, Venezuela; ⁵Ministry of Health, Mision Barrio Adentra, Puerto Ayacucho, Venezuela; ⁶Universidad de Ciencias Medicas, Mariana Grajales Coello, Holguin, Cuba; ⁷Department of Chemistry & Biochemistry, University of Colorado, Boulder, CO, USA; ⁸Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA, USA and ⁹Howard Hughes Medical Institute, Boulder, CO, USA

The human skin harbors complex bacterial communities. Prior studies showing high inter-individual variation focused on subjects from developed countries. We therefore compared cutaneous bacterial communities of Amerindians in the Venezuelan Amazon with subjects in the United States. Forearm skin specimens were studied from healthy Amerindians in Platanilla village in Amazonas State, and from healthy persons in New York and Colorado. All skin sampling used similar swab/buffer techniques. Multiplexed V2-targeted 16S rRNA gene pyrosequencing yielded high quality sequences from 112 samples. The results show 20 phyla, with three (Proteobacteria, Firmicutes, Actinobacteria) predominating. US residents and Venezuelan Amerindians had significantly different forearm skin bacterial community compositions, with United States dominated by *Propionibacterium*. Among the Amerindians, there was a deep split based on bacterial community membership, with 30 and 42 samples, respectively, falling into each of the two groups, not associated with age, gender, or body mass index. One Amerindian group had diversity similar to the United States, but was dominated by *Staphylococcus* rather than *Propionibacterium*. The other Amerindian group was significantly more diverse and even than the US or the other Amerindian group, and featured a broad range of Proteobacteria. The results provide evidence that ethnicity, lifestyle and/or geography are associated with the structure of human cutaneous bacterial communities.

The ISME Journal (2013) 7, 85–95; doi:10.1038/ismej.2012.81; published online 16 August 2012

Subject Category: microbe–microbe and microbe–host interactions

Keywords: microbiome; cutaneous; microbial diversity; human; genetics

Introduction

The development of tools for high throughput DNA sequencing and computational support for the vast data sets thus generated has overcome the limitation of previous culture-based microbiological studies of the human skin (Leyden *et al.*, 1981, 1987; Kearney *et al.*, 1984; Roth and James, 1988; Chiller *et al.*, 2001), and has renewed interest in the characterization of the human cutaneous

microbiome using culture-independent techniques (Dethlefsen *et al.*, 2007; Oakley *et al.*, 2008; Avila *et al.*, 2009; Hamady and Knight, 2009; Nasidze *et al.*, 2009; Peterson *et al.*, 2009; Turnbaugh *et al.*, 2009; Zhang *et al.*, 2009; Qin *et al.*, 2010). Investigators now are using molecular, DNA-based methodologies to assess the microbial populations present in and on human skin (Fredricks, 2001; Paulino *et al.*, 2006, 2008; Gao *et al.*, 2007, 2008, 2010; Fierer *et al.*, 2008, 2010; Grice *et al.*, 2008, 2009; Costello *et al.*, 2009; Grice and Segre, 2011). However, these studies were all performed in humans who live in developed countries, as does the large-scale National Institute of Health (NIH)-initiated Human Microbiome Project (Peterson *et al.*, 2009), that shares this focus on humans living lifestyles associated with postmodern socioeconomic

Correspondence: MJ Blaser, Departments of Medicine and Microbiology, New York University Langone Medical Center, 550 First Avenue, OBV A606, New York, NY 10016, USA.

E-mail: Martin.Blaser@nyumc.org

Received 19 December 2011; revised 3 May 2012; accepted 5 June 2012; published online 16 August 2012

development. The microbiota in normal persons in developed countries has likely been substantially altered by changes associated with socioeconomic development, including frequent showering, and use of soap, cosmetics and antibiotics (Blaser and Falkow, 2009). To better understand potentially alternative human cutaneous microbiota that could more closely resemble that with which we coevolved (Blaser, 2006), it would be useful to assess individuals whose exposures to modern-type lifestyles have been more limited.

Our long-standing studies of the microbiota of Amerindians in Amazonas State in Venezuela (Ghose *et al.*, 2002; Marini *et al.*, 2007; Dominguez-Bello *et al.*, 2010) enabled us to address several important questions. We sought to compare Amerindian and US samples by examining specimens obtained from the same cutaneous site using similar methods. We sought to describe the composition and uniformity of the sampled Amerindian cutaneous microbiome in a community of Amerindian subjects, to identify correlates with particular clustering patterns, and to determine whether Amerindian and US populations could be distinguished. For these studies, we sampled a single cutaneous site, the volar mid-forearm, as this site is easily identified and sampled, permits examining laterality of populations, is washed less than hands, and is subject to an extensive sequence database that is currently growing (Paulino *et al.*, 2006; Gao *et al.*, 2007, 2008, 2010; Paulino *et al.*, 2008; Costello *et al.*, 2009; Staudinger *et al.*, 2011).

Materials and methods

Study subjects

We obtained samples from 72 people in the village of Platanillal, ~20 km South of Puerto Ayacucho, the capital of Amazonas State, as described earlier (Marini *et al.*, 2007). At the time of sampling in October 2008, there was no electricity or running water and streets were not paved in Platanillal. The Guahibo were traditionally nomadic hunter-gatherers (Steward and Faron, 1959; Salzano and Callegari-Jacques, 1988). Since the 1970s, most have been settled in government-built houses in small villages led by a political captain. Present-day Guahibos retain their Amerind language (Greenberg, 1987), high levels of endogamy and a relatively sedentary lifestyle (Steward and Faron, 1959; Salzano and Callegari-Jacques, 1988). Their traditional diet is based on maize, cassava and fish, with an occasional wild game. Sixty-eight of the subjects of this study were born in Platanillal and four were born in nearby settlements (San Pablito, Acqua bianco). Sixty-nine subjects were of Guahibo (Jivi) ethnicity and three others were related (Jivi/mestizo, Piapoco/Jivi and Econeiwa). The subjects ranged in age from 2 months to 80 years (median 9 years). In total, we examined specimens from the

right forearms of 72 subjects, who volunteered to participate in the study. Informed consent was obtained from subjects or their guardians according to human experimentation guidelines of the Helsinki Declaration of 1975, as revised in 1983. In brief, written informed consent was obtained from all adult participants, or from their parents, in the case of young children. The experimental protocol was agreed upon by the village chief (captain) and the local medical doctor, had the support of the community, and was approved by the Venezuelan Institute of Scientific Research and the Institutional Review Board of the University of Puerto Rico–Rio Piedras.

For comparisons, we examined four adults in New York City (NYC), age 28–55 years; for two subjects, we compared specimens obtained 10 months apart. All had bilateral forearm sampling; in total, we examined 12 specimens from NYC. Sampling of the NYC subjects was under a protocol approved by the New York University Langone Medical Center (NYULMC IRB). As another reference group from the USA, we also included in the analysis the results from 28 samples previously obtained from healthy people in Boulder, CO, USA, as described and approved by the University of Colorado IRB (Costello *et al.*, 2009). These bilateral forearm samples were taken on two consecutive days from seven adults, age 31–36 years.

Intestinal protozoa and helminths

From 49 of the 72 subjects in Platanillal, fecal specimens had been obtained to assess for the presence of common intestinal protozoa and helminths. Analysis was by stool microscopy, as described (Marini *et al.*, 2007). As relatively high frequencies of carriage were observed for helminths (*Ascaris lumbricoides* (49%), *Trichuris trichura* (49%) and *Ankylostoma species* (16%)) we included data on the carriage of these organisms in the metadata. Subjects were scored for presence or absence of each of these three helminth species, and for the composite number present. A parallel analysis was done for the presence of five common protozoa, including *Entamoeba coli* (63%), *Endolimax nana* (47%), *Chilomastix mesnili* (35%), *Giardia lamblia* (22%) and *Iodamoeba butschlii* (18%).

Sample collection and DNA extraction

For Platanillal and New York residents, the volar aspect of the forearms were sampled as described in prior publications (Gao *et al.*, 2007, 2010; Grice *et al.*, 2009; Staudinger *et al.*, 2011) by applying a cotton pledget soaked in ST solution (0.15 M NaCl with 0.1% Tween 20). After swabbing for ~30 s, the head of each swab was aseptically cut from the handle, centrifuged for 10 min at 5900 g, and then removed. After the centrifugation, the contents of the tubes were frozen at –80 °C, until processing. DNA from the contents was extracted using the

Qiagen DNeasy method (Gao *et al.*, 2007). DNA was quantitated using the NANODROP-1000 (Wilmington, DE, USA) and samples were frozen until sequencing was performed. The amounts of DNA ($\text{ng } \mu\text{l}^{-1}$) extracted from the samples from the Amerindians (median 3.4 (1.7–5.4)) and the New Yorkers (3.1 (1.9–3.4)) were similar. Coloradans were swabbed in the same manner as for the Amerindians and New Yorkers, although the full length of the forearm was swabbed, as described (Costello *et al.*, 2009), and DNA was extracted directly from the frozen (-80°C) swabs using the MO BIO PowerSoil DNA Isolation kit (Carlsbad, CA, USA), as described in (Costello *et al.*, 2009).

Amplification and sequencing of 16S rRNA genes

Samples from all subjects were amplified and barcoded for multiplex pyrosequencing using primers targeted to the V2 region of the bacterial 16S rRNA gene under uniform PCR conditions, as described (Fierer *et al.*, 2008; Costello *et al.*, 2009; Hamady and Knight, 2009). In brief, the forward primer (5'-GCC TTG CCA GCC CGC TCA GTC AGA GTT TGA TCC TGG CTC AG-3') contained the 454 Life Sciences primer B sequence, the broad-range bacterial primer 27 F, and a two-base linker sequence ('TC'). The reverse primer (5'-GCC TCC CTC GCG CCA TCA GNN NNN NNN NNN NCA TGC TGC CTC CCG TAG GAG T-3') contained the 454 Life Sciences primer A sequence, a unique 12-nt error-correcting Golay barcode used to tag each PCR product (designated by the Ns), the broad-range bacterial primer 338R, and a 'CA' linker sequence between the barcode and the rRNA primer. Reactions were carried out in triplicate using $0.6 \mu\text{M}$ forward and reverse primers, $3 \mu\text{l}$ template DNA and 1X HotMasterMix (5 PRIME, Gaithersburg, MD, USA). Thermal cycling consisted of 94°C for 3 min, followed by 35 cycles of 94°C for 45 s, 50°C for 30 s, and 72°C for 90 s, with a final extension of 10 min at 72°C . Replicate amplicons were pooled and cleaned using the UltraClean-htp 96-well PCR Clean-up kit (MO BIO). Amplicon DNA concentrations were determined using the Quant-iT PicoGreen dsDNA reagent and kit (Invitrogen, Grand Island, NY, USA). Amplicons were then pooled in equimolar ratios and cleaned. Amplicon pyrosequencing was carried out using primer A on a 454 Life Sciences Genome Sequencer FLX instrument (Roche, Branford, CT, USA).

Sequence analysis

Sequences were processed using the QIIME software package (Caporaso *et al.*, 2010). Sequences were removed from the analysis if they were <200 or >350 nt in length, had a mean quality score <25 , contained ambiguous characters, contained a homopolymer run exceeding 12 nt, did not contain the primer sequence or contained an uncorrectable barcode. Remaining sequences were assigned to samples by examining the 12-nt barcode. The complete data set was error-corrected and chimera-

checked using OTUpipe (<http://drive5.com/otupipe>) (Caporaso *et al.*, 2010), which provided essentially the same results as Denoiser (Reeder and Knight, 2010) in terms of alpha diversity, beta diversity and taxonomic composition (data not shown). After this filtering process, data from residents of Platanilla, New York and Colorado were pooled, and similar sequences were clustered into operational taxonomic units (OTUs) using UCLUST (Edgar, 2010), with a minimum sequence identity threshold of 97%. In total, 1148 OTUs were obtained. A representative sequence was chosen from each OTU by selecting the first sequence (the UCLUST cluster seed). Taxonomy was assigned to each representative sequence using the Ribosomal Database Project (RDP) classifier (Wang *et al.*, 2007), with a minimum support threshold of 80% and the RDP taxonomic nomenclature. Representative sequences were aligned against the Greengenes core set using PyNASt software (Caporaso *et al.*, 2010), using a minimum alignment length of 150 and a minimum identity of 75%. From the total of 1148 OTUs, 1143 (99.5%) representative sequences aligned successfully and five that failed to align were dropped. The PH Lane mask was used to remove hypervariable regions after alignment. A phylogenetic tree was inferred using FastTree (Price *et al.*, 2009), with Kimura's 2-parameter model. For alpha diversity measurements, rarefaction analysis was performed for: (i), the estimated number of OTUs in each sample using the Chao1 algorithm; (ii) the evenness of the populations, using the Shannon measure of evenness and (iii), the amount of phylogenetic branch length observed in each sample (phylogenetic distance (PD)). For beta diversity measurements, UniFrac distances were calculated between all pairs of samples. UniFrac distances are based on the fraction of branch length shared between two communities in a phylogenetic tree. Unweighted UniFrac accounts for membership only; weighted UniFrac accounts for membership and relative abundance (Lozupone and Knight, 2005; Lozupone *et al.*, 2010). UniFrac-based jackknifed hierarchical clustering was performed using unweighted pair group method with arithmetic mean (UPGMA) in QIIME. Principal coordinates analysis (PCoA) was also performed on the UniFrac distance matrices, and visualized using the KiNG graphics program (<http://kinemage.biochem.duke.edu/software/king.php>).

Statistics

ANOVA on OTU and genus-level abundances was performed in QIIME. ANOSIM and PERMANOVA on unweighted and weighted UniFrac distance matrices were performed in PRIMER-E.

Data deposition

The bacterial 16S rRNA gene sequences reported in this work were deposited in MG-RAST.

Results

Global patterns

In total, sequencing yielded 31 765 high-quality reads from the 72 Platanillo and 12 New York forearm skin microbiota samples (mean 378 reads per sample with 229 nt mean read length) for comparison with the 27 290 sequences (mean 974 reads per sample with 229 nt mean read length), obtained from the 28 Colorado samples examined in a prior study (Costello *et al.*, 2009). Averaging across samples, Proteobacteria (33%), Firmicutes (29%) and Actinobacteria (27%) were the most abundant phyla of the 25 that were detected in total. *Staphylococcus* (22%) was the predominant genus, followed by *Propionibacterium* (11%), *Streptococcus* (3.1%), and *Pseudomonas* (3.1%) and *Corynebacterium* (2%). The paired specimens (left and right) obtained at two time points 10 months apart from New York showed highly similar patterns, as did the Colorado specimens obtained a day apart (Figure 1); samples 10 months apart showed greater variation, consistent with prior studies (Gao *et al.*, 2007).

Beta diversity and taxon representation patterns among US and Amerindian samples

UPGMA clustering and phylum-level abundances. We first used jackknifed hierarchical clustering (UPGMA), based on weighted UniFrac distances to investigate the relationships among the forearm skin bacterial communities. In an initial analysis, we examined 66 samples in which sampling yielded sufficient depth (>400 sequences per samples) to permit a well-supported dendrogram. By this analysis, there was a clear separation between the US (NYC and Colorado) and Amerindian specimens (Figure 1a). Within the Amerindian specimens, there was a deep branching distinguishing two clusters: Cluster A ($n=17$ samples) and Cluster B ($n=17$ samples). Clustering at lower coverage revealed a similar pattern, permitting assignment of all 72 Amerindian samples to either Cluster A ($n=30$) or B ($n=42$). Differences in DNA concentrations in the extracted samples do not account for the clustering into groups A or B (data not shown).

The predominant phyla in the US and the Venezuelan samples differed substantially, as illustrated in Figure 1b. The US specimens were dominated by Actinobacteria, irrespective of whether the subjects were from Colorado or from

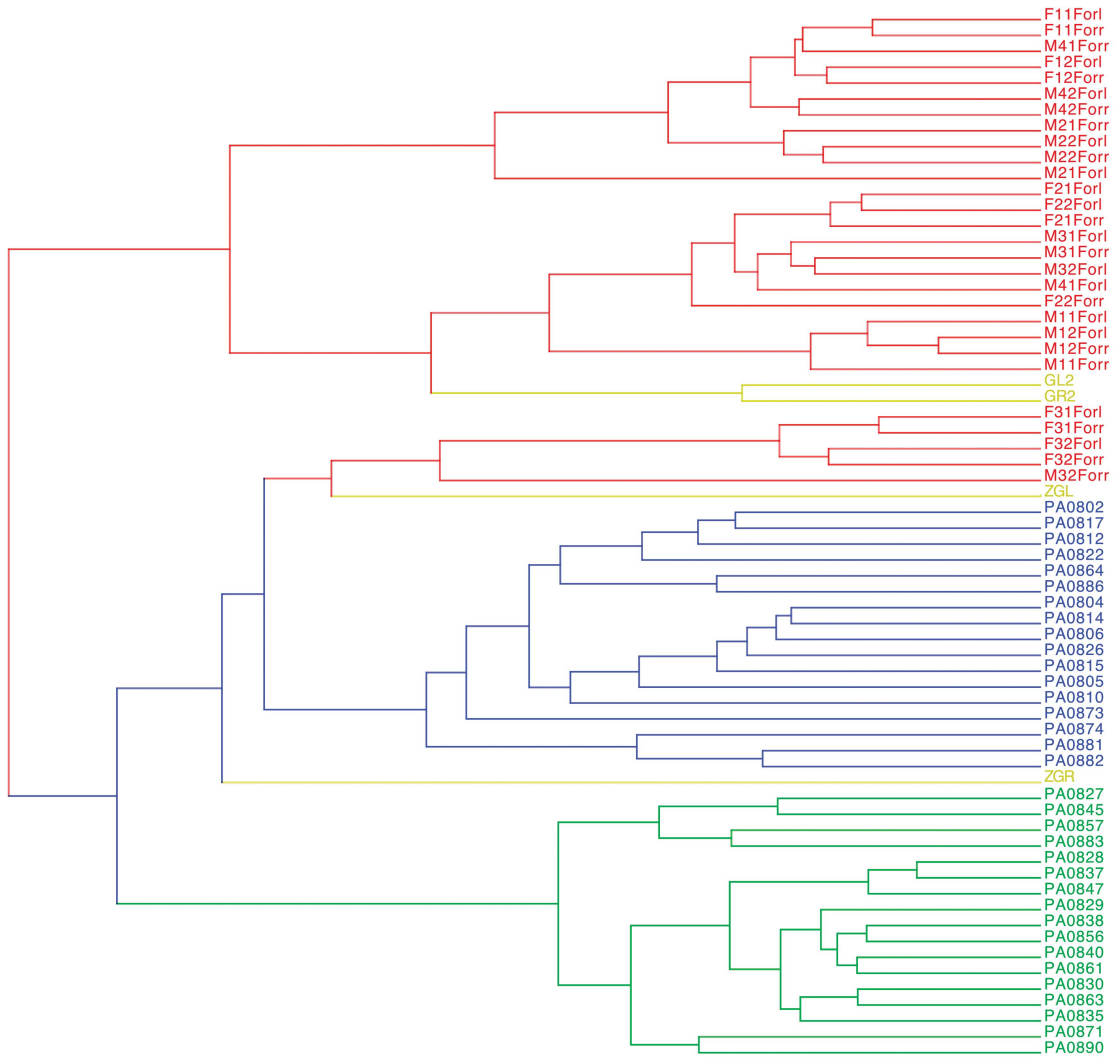
New York. That the taxa were similar between the two US groups indicates that differences in the extent of the forearm that was sampled and the difference in DNA extraction methods had little effect on the overall assessment of microbiome composition. In both Amerindian clusters, Actinobacteria comprised a substantially smaller fraction than in the US subjects. In contrast to the US samples, the Amerindian specimens were either dominated by Firmicutes (Cluster A), or showed no dominant phylum (Cluster B). There was higher representation of Bacteroidetes (9.51%), in Cluster B than in Cluster A (3.44%), or the US specimens (3.78%) ($P<0.001$ for each vs Cluster B). Thus, there are substantial compositional differences in the USA and Amerindian (Venezuela (VZ)) forearm microbiome.

Principal coordinates analyses and genus-level abundances. To better reflect the origins of the differences in Figure 1, we subjected the same UniFrac distance matrices to PCoA. After rarefying the data set to 400 sequences per sample, PCoA of the unweighted and weighted UniFrac distances (Figure 2) disclosed that the primary factor influencing differentiation of forearm skin bacterial community membership and structure (as shown in PC1) was country (US vs VZ; ANOSIM $P<0.001$; PERMANOVA $P<0.001$ for both unweighted and weighted measures). The US subjects from NY and CO had significantly higher average proportions of *Propionibacterium* (17% and 28%, respectively) compared with the VZ subjects (3%; ANOVA corrected for multiple comparisons $P<0.001$). This result was mainly driven by a single OTU (at the 0.97 level), whose representative sequence was confirmed by BLAST to be 100% identical to *Propionibacterium acnes*.

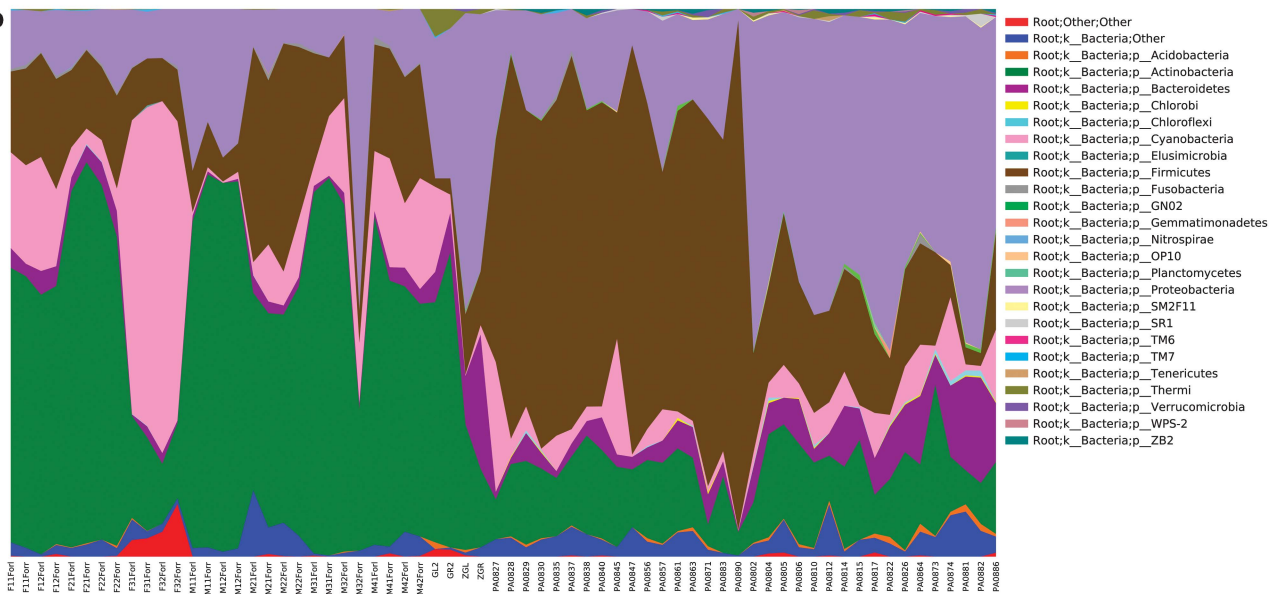
The UniFrac-based PCoA also revealed that the forearm skin bacterial assemblages from the Amerindians fell into two distinct clusters (along PC2) with roughly equal numbers of subjects in each group; this pattern was especially evident in the weighted UniFrac analysis, suggesting differentiation based on the relative abundances of particular taxa between the two groups, and consistent with the UPGMA analysis (Figure 1). On average, samples from Amerindian Cluster A had significantly higher proportions of the genus *Staphylococcus* compared with samples from Amerindian Cluster B and from

Figure 1 Diversity among cutaneous forearm samples. The 66 samples with >400 sequences are compared in this analysis. (a) UPGMA based on weighted UniFrac distances. The code for all Amerindian subjects begins with Puerto Ayacucho (PA). For the 28 samples from the CO subjects (red), M or F refers to male or female; l and r, refer to left and right, respectively and numbers refer to subject identity. For the four NY specimens (blue), L and R refer to left and right, respectively. Jackknife analysis of the weighted UniFrac UPGMA tree with 100 repetitions at 400 sequences per sample shows strong support for the distinction of US and Amerindians. Of the samples from the Amerindian subjects, a deep branching was identified, dividing the samples into Cluster A (purple) and Cluster B (green). Several of the US samples are clustered with Cluster B, but none with Cluster A. (b) Taxon summary at phylum level for the 66 samples. The 34 Amerindian subjects shown are divided based on whether they are in Cluster A (gold; $n=17$) or B (green; $n=17$). The predominant phyla are indicated by color (see key).

a



b



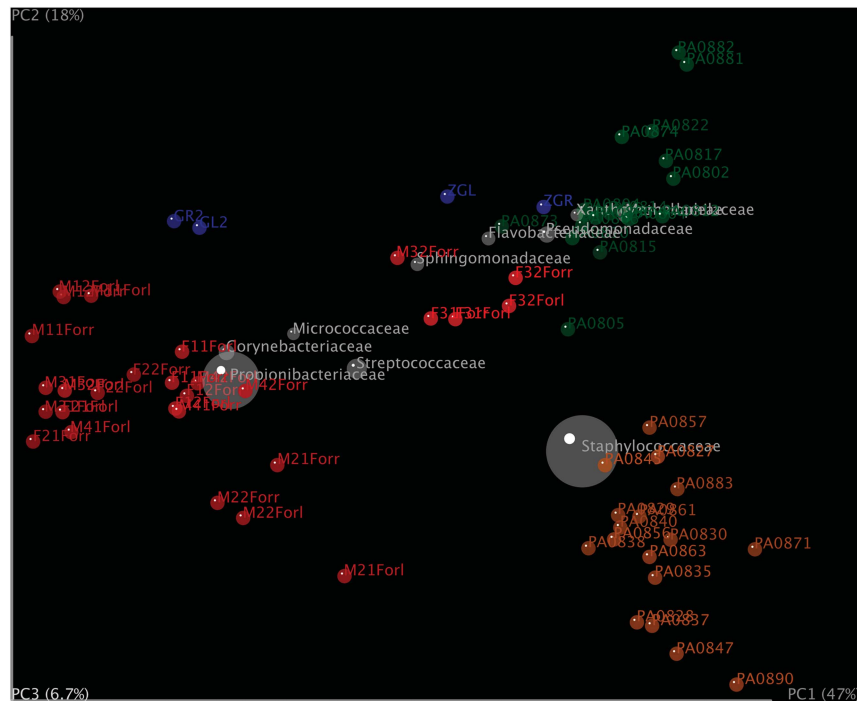


Figure 2 Clustering of study subjects using PCoA based on weighted UniFrac distances. Samples represented by coloring, by location and by clustering (based on Figure 1). Colors are: red (Colorado), blue (NYC), green (Amerindian cluster B), orange (Amerindian cluster A). In the Bi-plot, 10 predominant OTUs are indicated by the size of the gray circle representing the abundance of the taxon.

the US subjects (50% in A, 6.9% in B, 5.5% in NY, 4.4% in CO; ANOVA with correction for multiple comparison $P < 0.001$). This result was driven primarily by two OTUs, whose representative sequences were confirmed by BLAST to be 100% identical to *S. epidermidis* (35%) and *S. hominis* (12%). Indeed, every subject in Amerindian Cluster A had a relative abundance of the genus *Staphylococcus* $> 30\%$; this is consistent with the Firmicutes dominance observed in Figure 1b. Conversely, samples from the Amerindian Cluster B were far less dominated by a single genus such as *Propionibacterium* or *Staphylococcus*, and were instead characterized by a broad diversity of Proteobacteria, with the most abundant being OTUs classified as *Pseudomonas* (4.7%), *Xanthomonadaceae* (3.4%) and *Methylophilus* (2.7%). As such, overall diversity and evenness also were significantly higher in Cluster B than in the *Staphylococcus*-dominated Cluster A. The clustering was robust by location at sampling levels of 400 sequences per sample to 100 sequences per sample (Supplemental Figure 1). However, there was no effect on the clustering seen that could be attributed to subject gender (Supplemental Figure 2A), age (Supplemental Figure 2B) or body mass index (Supplemental Figure 2C).

Pairwise UniFrac distances. We next used the weighted UniFrac metric to examine pairwise distances between the samples according to their origin

and cluster. At the country level (Figure 3a), the smallest pairwise differences were observed for the Amerindian samples from Venezuela (VZ); intragroup distances for the US samples were significantly greater ($P < 0.001$) than for the VZ samples. The intergroup distance between the Venezuelan and US specimens was significantly greater ($P < 0.001$) compared with both of the intragroup distances (VZ–VZ and US–US).

Expanding the analysis, the lowest intragroup distances were within Venezuelan (Amerindian) Cluster A, followed by the Venezuelan Cluster B (Figure 3b). The mean pairwise distances between Clusters A and B were similar to that observed within the group of US subjects. Not surprisingly, the greatest ($P < 0.001$) pairwise distances were between Cluster A and the US sample, and between Cluster B and the US sample; these two metrics were essentially equivalent (Figure 3b). By PERMANOVA, in all pairwise comparisons, inter-cluster distances were significantly larger than intra-cluster distances. We also separated the US subjects into those from Colorado and New York. The distances between Colorado and New York were significantly lower than those for either group with either of the Venezuelan groups A and B (Supplemental Figure 3).

Effect of taxon dominance on the observed clustering. In total, 14 taxa with representation $> 1\%$ in the entire sample (common taxa) constituted 49.4% of the

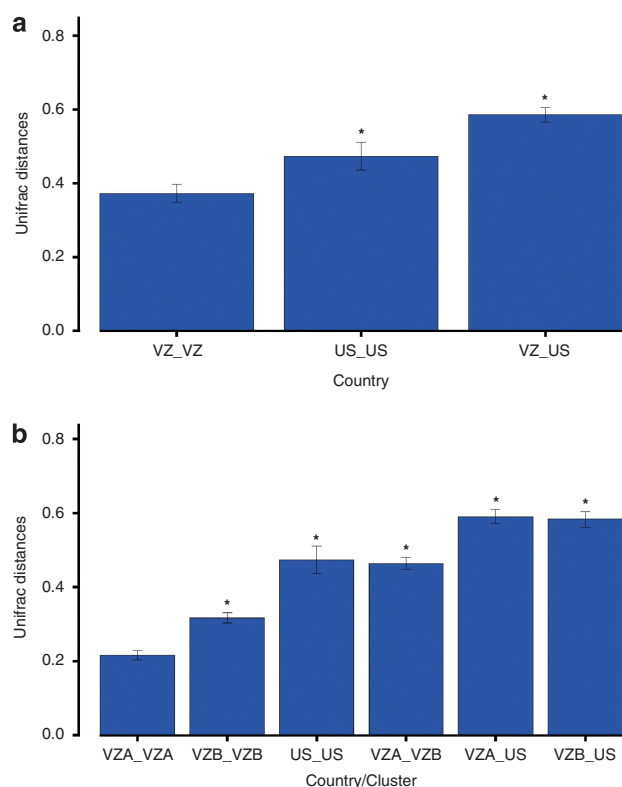


Figure 3 Weighted UniFrac pairwise distances between several groups of samples. Groups shown are from US, Venezuela (VZ) samples, and Amerindian Cluster A (VZA), Amerindian Cluster B (VZB). Intra-group and inter-group distances compared using the Permanova statistic, * $P < 0.001$ in comparison to the reference comparison, at left. (a) Intra- and inter-country distances. (b) Intra- and inter-cluster distances.

sequences. Another 113 taxa, with representation between 0.1% and 1.0% (intermediate taxa), constituted 29.6% of the sequences, and 1021 taxa with <0.1% representation (rare taxa) constituted 20.9% of the sequences. We examined whether the clustering that was observed primarily reflected clustering of the common (>1%), intermediate (0.1–1.0%) and/or rare (<0.1%) taxa. We used weighted (Figures 4a–c) and unweighted (Figures 4d–f) UniFrac analyses with PCoA representations to address this question. Because the numbers of sequences diminish with this stratification, we examined specimens with >100 sequences for each group. In each bin, whether weighted or not, the phylogenetic clustering observed in the total sample is present. Thus, the differences in US and Amerindian assemblages, including the A and B Clusters, are essentially uniform, involving taxa that are common, intermediate and rare. This consistency highlights the robustness of the differences observed between the US and Amerindian populations, and also indicates that the primary differences are not due to the different DNA extraction procedure used in the Colorado samples and the others. Procrustes analyses compared the original beta-diversity results vs the

common, medium and rare abundance OTUs (Supplemental Figure 4). Monte Carlo P values were significant ($P < 0.001$) for all comparisons using 1000 replicates. The split between USA and Amerindian holds at all levels, but the separation between Amerindian Clusters A and B only is notable using the common OTUs.

Measures of alpha diversity. Using the Chao1 algorithm for the samples at the level of 400 sequences per sample, the samples from New York and Colorado show essentially the same number of estimated species. Amerindian Cluster A has very similar α -diversity to the US samples, but Cluster B has significantly ($P < 0.05$) higher α -diversity (Figure 5a). This is consistent with the taxon distribution (shown in Figure 1b). In fact, the trends and differences observed when 400 sequences were evaluated from each sample also were present when as few as 100 sequences were studied. The Shannon measure of evenness shows the highest values for Cluster B (Figure 5b); again studying 100 sequences is sufficient to observe these trends. Alpha diversity, as measured by the amount of branch length in a phylogenetic tree (PD) (Figure 5c) was higher on average among samples from Cluster B, whereas Cluster A samples resembled the US samples. These results confirm the alpha diversity results shown in Figure 5a.

Village exposures. To explain the two distinct Clusters (A and B) of microbiota within residents of Platanillal, we explored exposure characteristics of the subjects. All Platanillal subjects reported drinking water from a well, and also using soap while washing. Most subjects reported bathing in water from wells, but 7 (11%) of the 65 subjects reported also bathing in the river; however, river-bathing rates in Clusters A (12%) and B (10%) were similar. Six (37.5%) of the 16 households sampled had both clusters represented among the subjects sampled (Supplemental Figure 4). Small households tended to have more Cluster B subjects. In 8 of 10 small (size ≤ 4) households, Cluster B persons were more common than Cluster A persons ($P = 0.002$). In the households in which only one cluster was found, Cluster B was represented in eight households ($P = 0.002$). There was no obvious relationship of clustering with the order in which households were sampled (Supplemental Figure 5).

Intestinal protozoa and helminths. Of the 49 Amerindian subjects sampled, 43 (88%) had ≥ 1 protozoan specimen detected (range 1–4), and 32 (65%) had ≥ 1 of the helminth species of interest (range 1–3). In total, 46 (93.9%) of the 49 subjects had ≥ 1 of the assayed protozoa and/or helminths detected by microscopy, with a mean of 3.0. Subjects in Cluster A trended toward greater parasite positivity rates for both protozoa and helminths, and

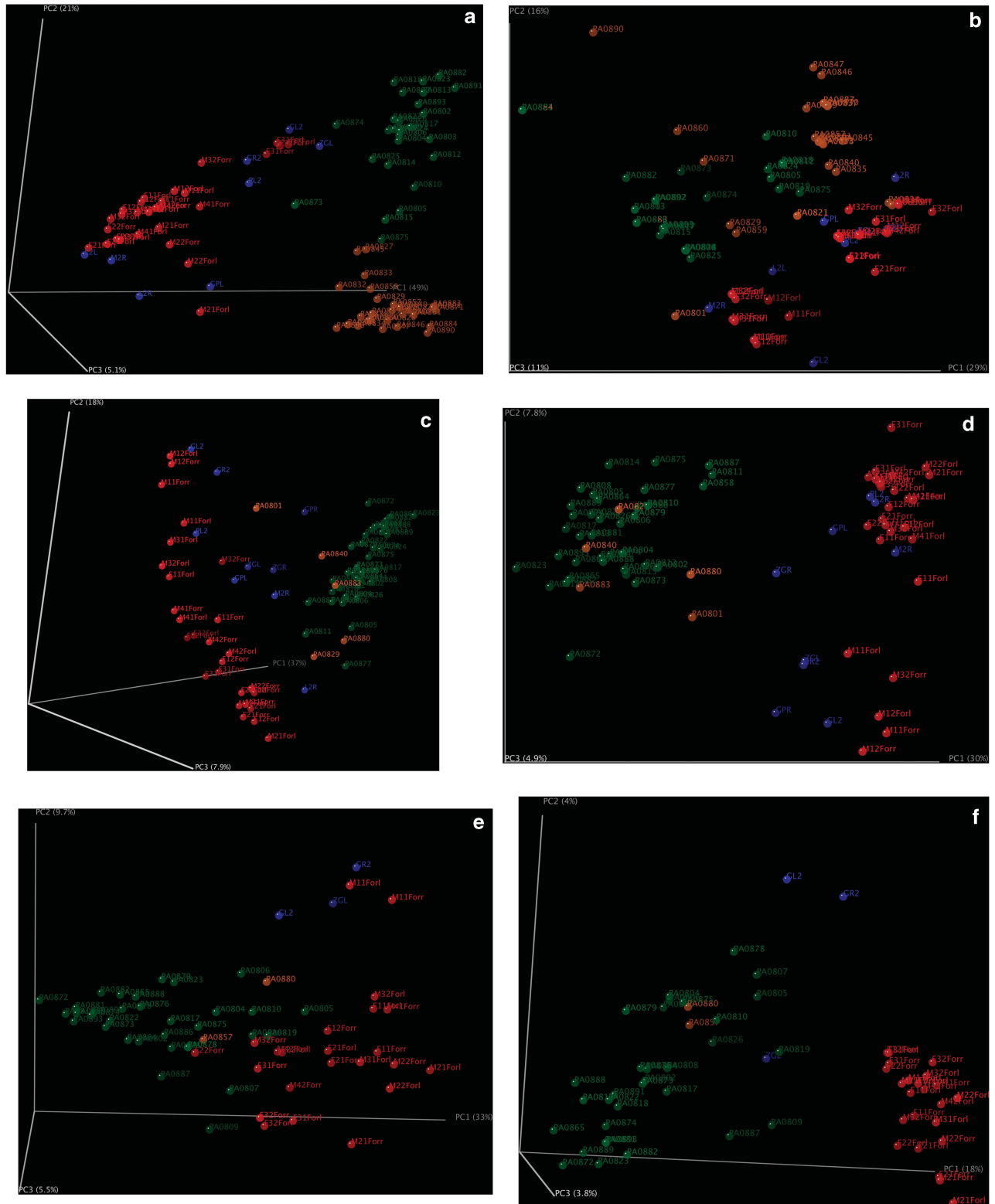


Figure 4 Effect of taxon abundance on clustering. The taxa represented in the 58 000 sequences were divided into three bins, and examined in weighted (a, c and e) and unweighted (b, d and f) UniFrac analyses. Taxa representing >1.0% (common) of the total abundance (a and b; 14 OTUs; 49.4% of all sequences); 0.1–1.0% (intermediate) of the total abundance (c and d; 113 OTUs; 29.6% of all sequences); and <0.1% (rare) of the total abundance (e and f; 1021 OTU; 20.9% of all sequences). Analyses were performed at the level of 100 sequences per sample; those samples with smaller numbers of sequences at particular stratifications are not shown. Clustering was compared using Permanova analysis to determine significant differences between US vs VZ, and Cluster A vs B. The colors are: Colorado (red); NYC (blue); Amerindian Cluster A (orange); Amerindian Cluster B (green).

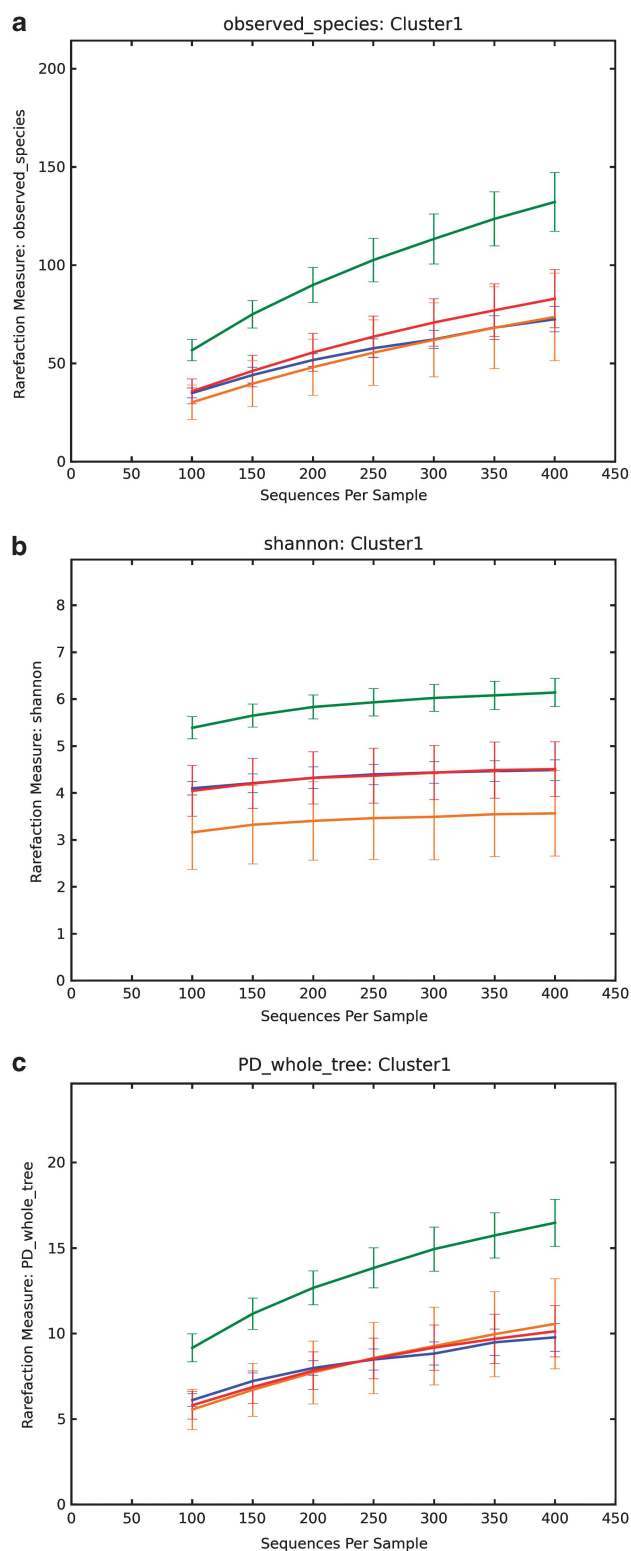


Figure 5 Measures of intra-group diversity. (a) Alpha diversity rarefaction using observed species in Chao1 analysis. Colors are: red (Colorado), blue (NYC), orange (Amerindian Cluster A), green (Amerindian Cluster B). Values represent mean + s.d. at each level of sampling. (b) Shannon score for evenness. (c) Phylogenetic distances.

had higher mean numbers of organisms detected (Supplemental Table 1), but differences between the two clusters were not significant.

Discussion

Our studies show clear differences in the cutaneous microbiota at a single reference site (forearm) between peoples in the United States and Amerindians in Venezuela. The similarities between the samples from the two US locations to one another in relation to the Amerindian samples (Supplemental Figure 1), despite some differences in study methods, are consistent with prior US studies of the cutaneous microbiota that show conservation of the major taxa (Gao *et al.*, 2007, 2010; Grice *et al.*, 2009; Costello *et al.*, 2009). In contrast, the microbiota from the Amerindians was substantially different from the United States, and was divided into two major clusters that were essentially not represented in the US samples. The alpha diversity indices indicate relative similarities in species richness (and evenness) among persons differing substantially in both ethnicity and life style (United States and Cluster A); however, for a second group of Amerindian samples (Cluster B), there was much higher richness and evenness. Thus, Cluster B differed from both the United States and Cluster A, and the differences were apparent even at low sampling depth (100 sequences or sample; Figure 5).

Both the US sample and Cluster A were dominated by a single major taxon, but which differed from one another (*Propionibacterium* in the US sample, and *Staphylococcus* in Cluster A), whereas in Cluster B, there was no single dominant taxon. These observations bear some resemblance to other human body sites, for example, the enterotypes described in the intestinal microbiome (Arumugan *et al.*, 2011), or the different major patterns observed in the vagina (Zhou *et al.*, 2010), in that there are clusters with continuous variation linking the clusters. However, in these latter cases, the major types are present in all geographic groups studied, whereas Cluster B was found essentially only in the Amerindians. However, sampling of other populations will be needed to determine the diversity of clustering patterns.

An alternative possibility is that the populations clustered are not fixed, but could change from one predominant type into the other, such as has been sometimes observed in longitudinal studies of the vaginal microbiota (Ravel *et al.*, 2011). Nevertheless, although the US sample and Cluster A have a dominant taxon, the differences between the three groups (US, Cluster A, and Cluster B) extend to taxa that have intermediate or rare representation (Figure 4). Thus, at this single point of sampling, the patterns are robust. This also is shown by the UniFrac analysis of the pairwise distances, in which the

inter-group distances always are greater than the intra-group distances. These findings indicate that on average, when the entire microbial community is taken into account (rather than just the first few principal components), individuals from the same group are more similar than individuals in different groups.

Platanillal, the village sampled is representative of a population in transition; two to three generations ago its peoples were nomadic hunter-gatherers (Steward and Faron, 1959; Salzano and Callegari-Jacques, 1988), but now the peoples are settled in permanent homes, with access to aspects of modern life, yet with a still relatively traditional diet. Reflecting a lack of piped water, carriage of intestinal protozoa and helminths is universal (Supplemental Table 1), and multiple species is the rule. Nevertheless, although the hypothesis that carriage of parasites (helminths and protozoa) changes immunological balance (for example, in a Th2 direction) (Maizels *et al.*, 2009; Fyhrquist *et al.*, 2012), which could then affect microbiota composition, is attractive, we did not find support for this idea, and we could not identify a factor that distinguished between carriage of a Cluster-A or Cluster-B cutaneotype. There was no association with age, gender, body mass index, relation to bathing, drinking water, or use of soap, or parasite type or overall burden and individuals differing in type often lived in the same household. Further sampling and longitudinal studies will be needed to discern the important correlates.

Finally, there is intensive interest in variation of the human microbiome in relation to health and disease (Peterson *et al.*, 2009; Qin *et al.*, 2010). However, many of the constituents of the microbiota have deep ancestral relationships with their hosts (Linz *et al.*, 2007; Ley *et al.*, 2008). Modernization is changing our microecology (Blaser, 2006; Dethlefsen *et al.*, 2008; Blaser and Falkow, 2009). To understand the composition and function of the human microbiome before the development of processes associated with postmodern life, we should seek and sample indigenous people before they become more subject to modernization, which is rapidly advancing around the world. By sampling more groups of individuals in varied locales who have retained more traditional lifestyles, we may be better able to discover more extensive variation in the cutaneous microbiota.

Acknowledgements

This work was supported in part by the Diane Belfer Program for Human Microbial Ecology, Biomedical Research Foundation of Colorado, SACAICET, Venezuela and by NIH grants HG004872, HG004866, DK090989 and UH2 AR057056 from the Human Microbiome Project, and by the Howard Hughes Medical Institute. We thank Donna Berg-Lyons for her technical assistance and Dr Huilin Li for her statistical support.

References

- Arumugan M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR *et al.* (2011). Enterotypes of the human gut microbiome. *Nature* **473**: 174–180.
- Avila M, Ojcius DM, Yilmaz O. (2009). The oral microbiota: living with a permanent guest. *DNA Cell Biol* **28**: 405–411.
- Blaser MJ. (2006). Who are we? Indigenous microbes and the ecology of human diseases. *EMBO Reports* **7**: 956–960.
- Blaser MJ, Falkow S. (2009). What are the consequences of the disappearing human microbiota? *Nat Rev Microbiol* **7**: 887–894.
- Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. (2010). PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26**: 266–267.
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335–336.
- Chiller K, Selkin BA, Murakawa GJ. (2001). Skin microflora and bacterial infections of the skin. *J Invest Dermatol Symp Proc* **6**: 170–174.
- Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. (2009). Bacterial community variation in human body habitats across space and time. *Science* **326**: 1694–1697.
- Dethlefsen L, Huse S, Sogin ML, Relman DA. (2008). The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol* **6**: 280.
- Dethlefsen L, McFall-Ngai M, Relman DA. (2007). An ecological and evolutionary perspective on human–microbe mutualism and disease. *Nature* **449**: 811–818.
- Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N *et al.* (2010). Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci USA* **107**: 11971–11975.
- Edgar RC. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.
- Fierer N, Hamady M, Lauber CL, Knight R. (2008). The influence of sex, handedness, and washing on the diversity of hand surface bacteria. *Proc Natl Acad Sci USA* **105**: 17994–17999.
- Fierer N, Lauber CL, Zhou N, McDonald D, Costello EK, Knight R. (2010). Forensic identification using skin bacterial communities. *Proc Natl Acad Sci USA* **107**: 6477–6481.
- Fredricks DN. (2001). Microbial Ecology of Human Skin in Health and Disease. *J Invest Dermatol Symp Proc* **6**: 167–169.
- Fyhrquist N, Lehtimäki S, Lahl K, Savinko T, Lappeteläinen AM, Sparwasser T *et al.* (2012). Foxp3+ cells control Th2 responses in a murine model of atopic dermatitis. *J Invest Dermatol* **132**: 1672–1680.
- Gao Z, Perez-Perez GI, Chen Y, Blaser MJ. (2010). Quantitation of major human cutaneous bacterial and fungal populations. *J Clin Microbiol* **48**: 3575–3581.
- Gao Z, Tseng CH, Pei Z, Blaser MJ. (2007). Molecular analysis of human forearm superficial skin bacterial biota. *Proc Natl Acad Sci USA* **104**: 2927–2932.

- Gao Z, Tseng CH, Strober B, Pei Z, Blaser MJ. (2008). Substantial alterations of the cutaneous bacterial biota in psoriatic lesions. *PLoS ONE* **3**: 2719.
- Ghose C, Perez-Perez GI, Dominguez-Bello MG, Pride DT, Bravi CM, Blaser MJ. (2002). East Asian genotypes of *Helicobacter pylori*: strains in Amerindians provide evidence for its ancient human carriage. *Proc Natl Acad Sci USA* **99**: 15107–15111.
- Greenberg JH. (1987). *Languages in the Americas*. Stanford University Press: Stanford, USA.
- Grice E, Segre J. (2011). The skin microbiome. *Nat Rev Microbiol* **9**: 244–253.
- Grice EA, Kong HH, Conlan S, Deming CB, Davis J, Young AC *et al*. (2009). Topographical and temporal diversity of the human skin microbiome. *Science* **324**: 1190–1192.
- Grice EA, Kong HH, Renaud G, Young AC, Bouffard GG, Blakesley RW *et al*. (2008). A diversity profile of the human skin microbiota. *Genome Res* **18**: 1043–1050.
- Hamady M, Knight R. (2009). Microbial community profiling for human microbiome projects: tools, techniques, and challenges. *Genome Res* **19**: 1141–1152.
- Kearney JN, Harnby D, Gowland G, Holland KT. (1984). The follicular distribution and abundance of resident bacteria on human skin. *J Gen Microbiol* **130**: 797–801.
- Ley RE, Lozupone CA, Hamady M, Knight R, Gordon JL. (2008). Worlds within worlds: evolution of the vertebrate gut microbiota. *Nat Rev Microbiol* **6**: 776–788.
- Leyden JJ, McGinley KJ, Holzle E, Labows JN, Kligman AM. (1981). The microbiology of the human axilla and its relationship to axillary odour. *J Invest Dermatol* **77**: 413–416.
- Leyden JJ, McGinley KJ, Nordstrom KM, Webster GF. (1987). Skin microflora. *J Invest Dermatol* **88**: 65s–72s.
- Linz B, Balloux F, Moodley Y, Manica A, Liu H, Roumagnac P *et al*. (2007). An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature* **445**: 915–918.
- Lozupone C, Knight R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* **71**: 8228–8235.
- Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. (2010). UniFrac: an effective distance metric for microbial community comparison. *ISME J* **5**: 169–172.
- Maizels RM, Pearce EJ, Artis D, Yazdanbakhsh M, Wynn TA. (2009). Regulation of pathogenesis and immunity in helminth infections. *J Exp Med* **206**: 2059–2066.
- Marini E, Maldonado A, Cabras S, Hidalgo G, Buffa R, Marin A *et al*. (2007). *Helicobacter pylori* and intestinal parasites are not detrimental to the nutritional status of Amerindians. *Am J Trop Med Hyg* **76**: 534–540.
- Nasidze I, Li J, Quinque D, Tang K, Stoneking M. (2009). Global diversity in the human salivary microbiome. *Genome Res* **19**: 636–643.
- Oakley BB, Fiedler TL, Marrazzo JM, Fredricks DN. (2008). The diversity of human vaginal bacterial communities and their association with clinically defined bacterial vaginosis. *Appl. Environ. Microbiol* **74**: 4898–4909.
- Paulino L, Tseng CH, Blaser MJ. (2008). Analysis of *Malassezia* microbiota in healthy superficial human skin and in psoriatic lesions by multiplex real-time PCR. *FEMS Yeast Research* **8**: 460–471.
- Paulino L, Tseng CH, Strober B, Blaser MJ. (2006). Molecular analysis of fungal microbiota in samples from healthy human skin and psoriatic lesions. *J Clin Microbiol* **44**: 2933–2941.
- Peterson J, Garges S, Giovanni M, McInnes P, Wang L, Schloss JA *et al*. (2009). The NIH Human Microbiome Project. *Genome Res* **19**: 2317–2323.
- Price MN, Dehal PS, Arkin AP. (2009). FastTree: computing large minimum-evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* **26**: 1641–1650.
- Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C *et al*. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**: 59–65.
- Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, McCulle SL *et al*. (2011). Vaginal microbiome of reproductive-age women. *Proc Natl Acad Sci USA* **108** Suppl 1 4680–4687.
- Reeder J, Knight R. (2010). Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. *Nat Methods* **7**: 668–669.
- Roth RR, James WD. (1988). Microbial ecology of the skin. *Annu Rev Microbiol* **42**: 441–464.
- Salzano FM, Callegari-Jacques SM. (1988). *South American Indians. A Case Study in Evolution*. Clarendon Press: Oxford, UK.
- Staudinger T, Pipal A, Redl B. (2011). Molecular analysis of the prevalent microbiota of human male and female forehead skin compared to forearm skin and the influence of make-up. *J Appl Microbiol* **110**: 1381–1389.
- Steward JH, Faron LC. (1959). *Native Peoples of South America*. McGraw-Hill: New York.
- Turnbaugh PJ, Hamady M, Yatsunencko T, Cantarel BL, Duncan A, Ley RE *et al*. (2009). A core gut microbiome in obese and lean twins. *Nature* **457**: 480–484.
- Wang Q, Garrity GM, Tiedje JM, Cole JR. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**: 5261–5267.
- Zhang H, DiBaise JK, Zuccolo A, Kudrna D, Braidotti M, Yeisoo Y *et al*. (2009). Human gut microbiota in obesity and after gastric bypass. *Proc Natl Acad Sci U S A* **106**: 2365–2370.
- Zhou X, Hansmann MA, Davis CC, Suzuki H, Brown CJ, Schutte U *et al*. (2010). The vaginal bacterial communities of Japanese women resemble those of women in other racial groups. *FEMS Immunol Med Microbiol* **58**: 169–181.

Supplementary Information accompanies the paper on The ISME Journal website (<http://www.nature.com/ismej>)