

Cell Reports

Supplemental Information

**The Gut Microbiota of Rural Papua New Guineans:
Composition, Diversity Patterns,
and Ecological Processes**

Inés Martínez, James C. Stegen, Maria X. Maldonado-Gómez, A. Murat Eren, Peter M. Siba, Andrew R. Greenhill, and Jens Walter

Supplemental Information. “The gut microbiota of rural Papua New Guineans: Composition, diversity patterns, and ecological processes”. Martínez et al.

PNG lifestyle

Study participants included individuals from two traditional societies of PNG, the Asaro and the Sausi. The Asaro Valley is located in the highlands of PNG (Figure 1A), approximately 1,500 meters above sea level. Our Asaro study site was approximately a 30 min drive from the provincial capital of Goroka in Eastern Highlands province, which is serviced by air and road transport. Asaro is one of the larger ethnic groups in PNG, with approximately 50,000 people. In contrast, the Sausi live in the lowlands of PNG in the Ramu Valley, Madang province, and consist of approximately 1,000 people. Our Sausi study site of Nauna village is approximately 230 meters above sea level at the foothills of the Finisterre Ranges. Madang province is contiguous to Eastern Highlands but geographically distinct. Asaro and Nauna are 45 km apart (Figure 1A) and are joined by pedestrian tracks (2-3 days walk in steep terrain) and by road (235 km; 4-6 hours travelling time). Contact between the villages does occur, but is infrequent.

Thirty-eight out of the 40 individuals in both villages had occupations that are hardly economically remunerated, or not at all (such as housemaid work), and therefore had no reliable source of income and lived austere, traditional lifestyles. The highest reported level of education achieved was grades 1-6 for Sausi participants, while 25% had no formal education. In the Asaro village, two-thirds of the participants reportedly had no formal education, with the highest reported level of education being grades 7-10. The typical PNG household is comprised by an average of 6 individuals, although it can reach 10 to 12. Close contact with farm animals (i.e. pigs and chickens) is customary, although they are rarely a part of the household. A high proportion of births are unattended with deliveries occurring either at home or outdoors. Infants are almost

exclusively breastfed until 6 months of age, though incorporation of solid foods can occur from as early as 3 months of age. Breast milk generally complements their nutrition for at least the first two years of life. Persistent under-nutrition is rare in PNG. A kinship system ('Wontok') ensures food security. Nonetheless, caloric intake has been estimated at < 2,000 daily kcal/person for 42% of the population and appropriate micronutrient intake can be a concern (Gibson, 2000). Carbohydrate sources are readily available and constitute the majority of rural PNG diet. The staple foods are sweet potato, taro, and plantain. Other dietary sources of complex carbohydrates include banana, cassava, yam, and sago. Leafy green vegetables, fruits and legumes are commonly consumed. Antibiotic use is common in PNG and is poorly regulated. Based on self-reported data collected from individuals at the time of their admission to health care facilities unregulated antibiotic use in febrile individuals in the Asaro region is approximately 10% (our unpublished data); no such data exist for the Sausi.

Are PNG exclusive OTUs detected in other non-industrialized cohorts?

We investigated whether the core OTUs exclusively detected in PNG could represent specific lineages of the non-westernized fecal microbiota. Thus, we examined whether related lineages could be detected in the Hadza hunter-gatherers (Schnorr et al., 2014), Malawi and Amerindian (Yatsunenko et al., 2012), and children from Burkina Faso (De Filippo et al., 2010). Full-length 16S rRNA reads with 100% identical V5-V6 regions to the representative sequences of the *Lactobacillus reuteri*-like and *Helicobacter macacae*-like OTUs were retrieved from the NCBI database, and a BLAST search was conducted on the non-industrialized datasets using these full-length 16S rRNA sequences. No good matches for either phylotypes were found in any of these datasets. However, we detected 16S rRNA gene tags closely related to *Helicobacter cinaedi* in Malawians and Venezuelan-Amerindians, *Helicobacter pollorum* in the Malawian cohort,

and *Helicobacter* sp. in the Hadza samples, albeit in low abundances and only a small number of individuals (Suppl. Table S5). We did not detect *Helicobacter* sequences in the Burkina Faso cohort (Suppl. Table S5), but this dataset has lower sequencing depth than the others considered in this analysis. *Lactobacillus* populations in non-industrialized societies (Hadza, Malawi and Amerindian) were detected in very low abundance, and were present in their westernized counterparts. Only the de Filippo et al. (2010) (De Filippo et al., 2010) dataset contained lactobacilli OTUs exclusive to the African cohort, and they were closely related to *Lactobacillus fermentum*, *Lactobacillus ruminis*, and *Lactobacillus salivarius* (Suppl. Table S5). Therefore, it appears that, although *Helicobacter* and *Lactobacillus* lineages were exclusively detected in other datasets of humans living in non-industrialized societies, they consist of species other than the ones exclusively associated with the PNG cohort. The representative sequences of the 39 OTUs related to *Streptococcus lutetiensis/infantarius* and *Streptococcus equinus* exclusive to PNG participants had no identical matches in available databases (Greengenes, NCBI, RDP); thus, their presence in other datasets could not be tested.

***Streptococcus* as an important component of the PNG microbiota**

A notable finding in this study was the elevated proportion of streptococci in PNG fecal samples (> 20% on average) as well as their phylogenetic diversity. The streptococci detected in PNG were related to the species *S. lutetiensis/infantarius*, *S. gordonii*, *S. equinus*, *S. salivarius/thermophilus*, *S. gallolyticus*, *S. pyogenes*, *S. intestinalis*, while US individuals were dominated by only one OTU related to *S. salivarius/thermophilus*. To our knowledge, fecal streptococci have not been reported to be impacted by diet; but small intestinal *Streptococcus equinus*-, *lutetiensis*-, and *salivarius*-type isolates can ferment several mono-, di-, and tri- saccharides, and *Streptococcus equinus*- and

lutetiensis-type isolates ferment starch and cellobiose (van den Bogert et al., 2013) in vitro. Environmental factors other than diet are bound to contribute to the compositional variations observed between westernized and non-industrialized societies, such as hygiene, which likely impacts dispersal of microbes. *Streptococcus* abundance in fecal samples has been shown to be significantly higher in Bangladeshi children living in an urban slum when compared to US children (Lin et al., 2013), and has been associated with diarrhea in children in low-income countries (Pop et al., 2014). Although the relevance of fecal streptococci in human disease is unknown, streptococci are among the most relevant bacterial pathogens in PNG, due to their aetiological role in pneumonia and meningitis (Riley, 2009). Colonization by streptococci is common, with nasopharyngeal carriage of *Streptococcus pneumoniae* in PNG infants among the highest in the world (Aho et al., 2010). The high incidence of diarrhea in children of the Asaro Valley (Wyrsh et al., 1998) and the now observed high prevalence of streptococci in fecal samples of PNG individuals, in conjunction with the findings of Pop et al. (Pop et al., 2014), warrants further investigation of this bacterial genus regarding its involvement in pathogenicity in PNG.

Implications of alterations in westernized microbiomes for human health

It is exciting to speculate that dissimilarities between the westernized and non-westernized microbiota might contribute to differences in the incidence of chronic western illnesses (Noverr and Huffnagle, 2005). Because the gut microbiota shapes human metabolic phenotypes and immune responses, health consequences could arise from contraction/loss of beneficial taxa or the expansion of detrimental taxa, or from an overall alteration of the microbial community in that it cannot provide health-associated ecosystem services. Diversity is hereby considered an important attribute of ecosystems, and the reduction in microbiome diversity observed in westernized societies

is likely to be detrimental as it is also associated with many human pathologies (Webb and Forman, 1995). *Helicobacter* species have been proposed to provide metabolic and immunological benefits to humans (Blaser and Falkow, 2009), and *Lactobacillus reuteri* provides benefits when administered as a probiotic (as reviewed in Walter et al. (Walter et al., 2011)). On the other hand, the genus *Bilophila* enriched in the US cohort, has been linked to inflammation in the gut (Devkota et al., 2012). Clearly, comparisons between the westernized and non-westernized microbiomes cannot be used to determine causal relationships, but they can serve as the basis of hypotheses for clinical research targeting the gut microbiota to improve health. Although certain core bacterial lineages in PNG were completely undetected in the US, the large overlap in OTU membership across cohorts is encouraging in that lifestyle and dietary changes could potentially rectify the impact of westernization. Diets enriched in fruit, vegetables, or whole grains (David et al., 2014; Martínez et al., 2013) increase diversity and induce community features more similar to those observed in non-westernized samples and serve as testimony that such changes could potentially be achieved, and community members that are less prevalent in westernized cohorts could be restored as live biotherapeutics. In fact, fecal bacteriotherapy is in essence an attempt to optimize bacterial transmission from one person to another, and its success in treating *Clostridium difficile* infection can serve as a paradigm on how increasing bacterial dispersal can redress dysbiotic microbiomes to improve human health (van Nood et al., 2013).

Supplemental References

Aho, C., Greenhill, A., Phuanukoonnon, S., Michael, A., Moberley, S., Pomat, W., Siba, P., Richmond, P., Lehmann, D., and Neonatal, P.T.T. (2010). Impact of neonatal and early infant pneumococcal conjugate vaccination on pneumococcal carriage and

suppurative otitis media in Papua New Guinea. In International Symposium on Pneumococci and Pneumococcal Diseases, (Tel Aviv, Israel).

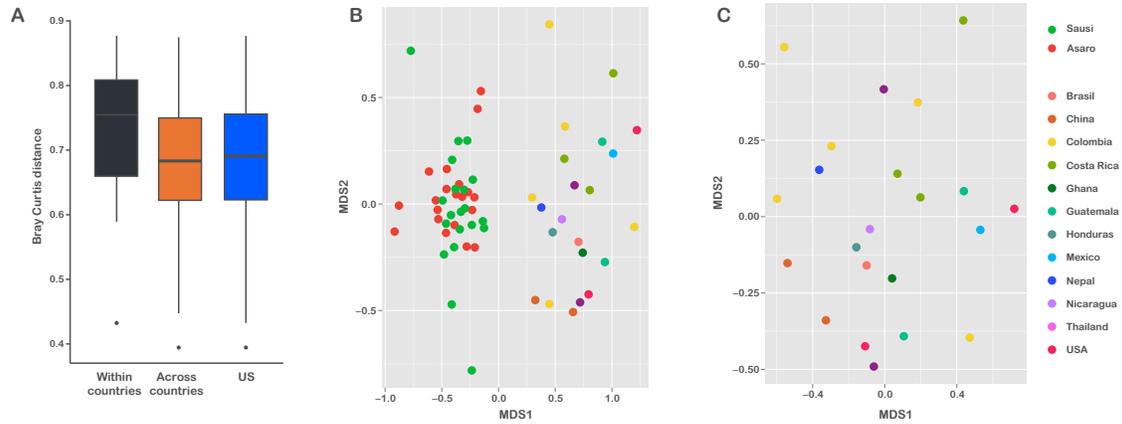
Van den Bogert, B., Erkus, O., Boekhorst, J., de Goffau, M., Smid, E.J., Zoetendal, E.G., and Kleerebezem, M. (2013). Diversity of human small intestinal *Streptococcus* and *Veillonella* populations. *FEMS Microbiol. Ecol.* *85*, 376–388.

Gibson, J. (2000). Nutritional Status of Papua New Guinea's Population and Its Determinants. In *Papua New Guinea Food and Nutrition*.

Lin, A., Bik, E.M., Costello, E.K., Dethlefsen, L., Haque, R., Relman, D.A., and Singh, U. (2013). Distinct distal gut microbiome diversity and composition in healthy children from Bangladesh and the United States. *PLoS One* *8*, e53838.

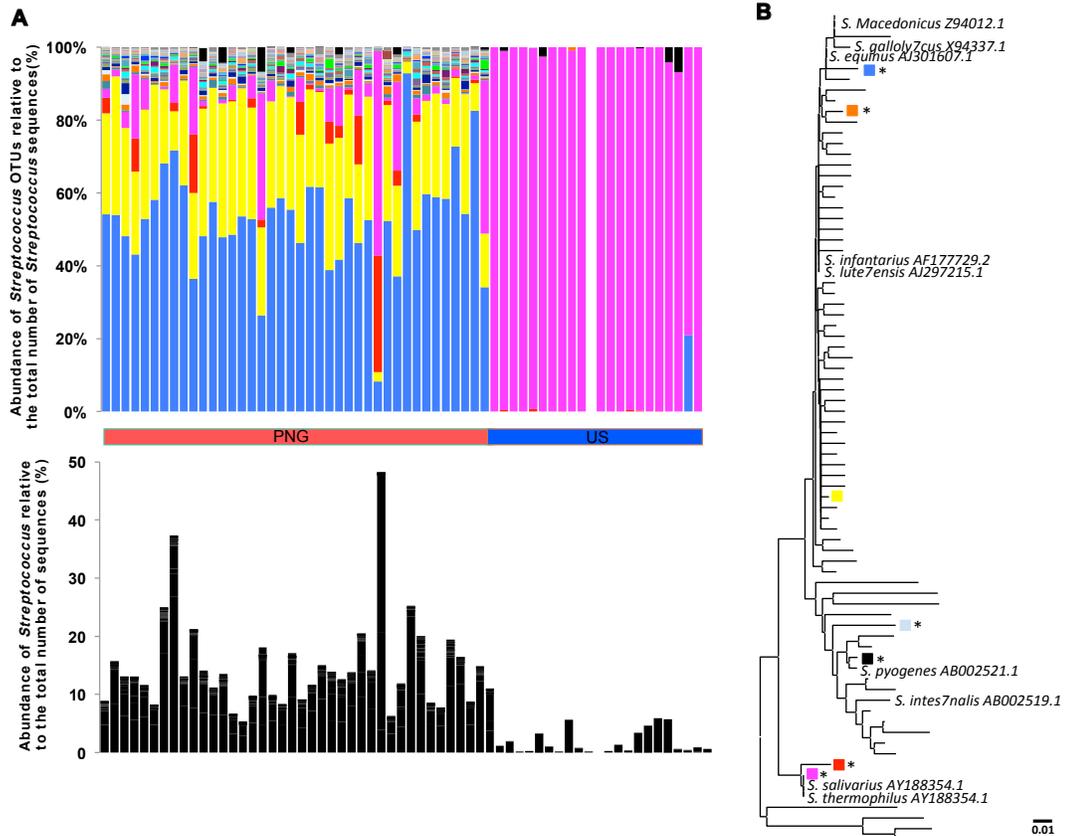
Van Nood, E., Vrieze, A., Nieuwdorp, M., Fuentes, S., Zoetendal, E.G., de Vos, W.M., Visser, C.E., Kuijper, E.J., Bartelsman, J.F.W.M., Tijssen, J.G.P., et al. (2013). Duodenal infusion of donor feces for recurrent *Clostridium difficile*. *N. Engl. J. Med.* *368*, 407–415.

Supplemental Figure S1 - Related to Figure 2.



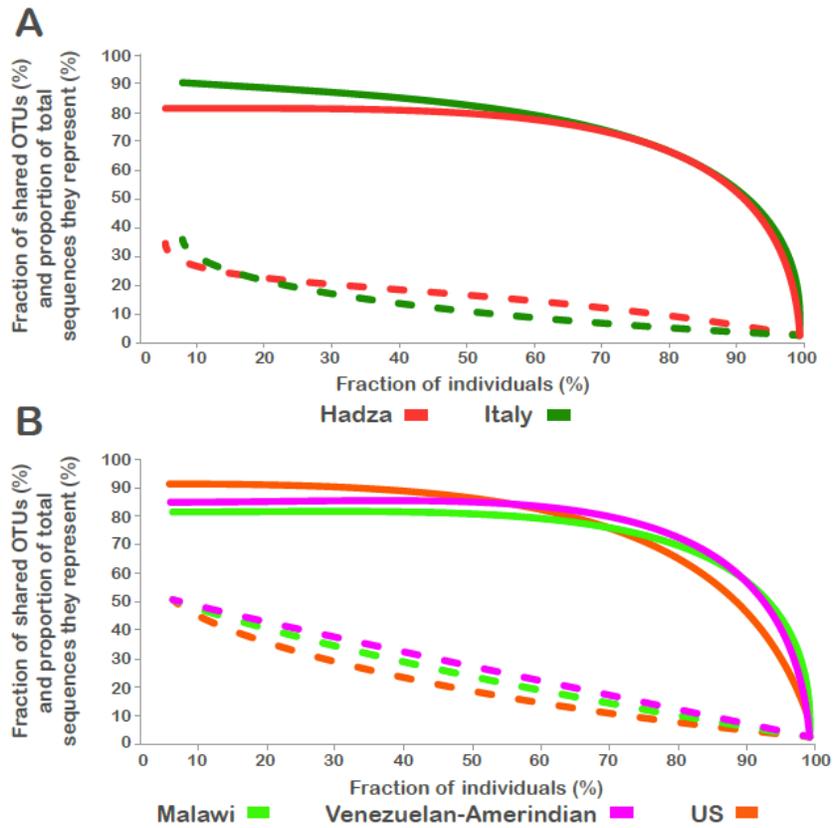
Supplemental Figure S1 - Related to Figure 2. Bacterial beta-diversity of US participants. Beta-diversity of the bacterial communities computed with Bray-Curtis diversity indices for subjects in the US cohort of the same country of origin (within countries), of different countries of origin (across countries), and for the entire US cohort (US) (A). NMDS ordination plots of fecal bacterial communities based on the Bray-Curtis distance metric for the PNG and US cohorts (B), and US cohort alone (C), color-coded by country of origin.

Supplemental Figure S2 - Related to Figure 3.



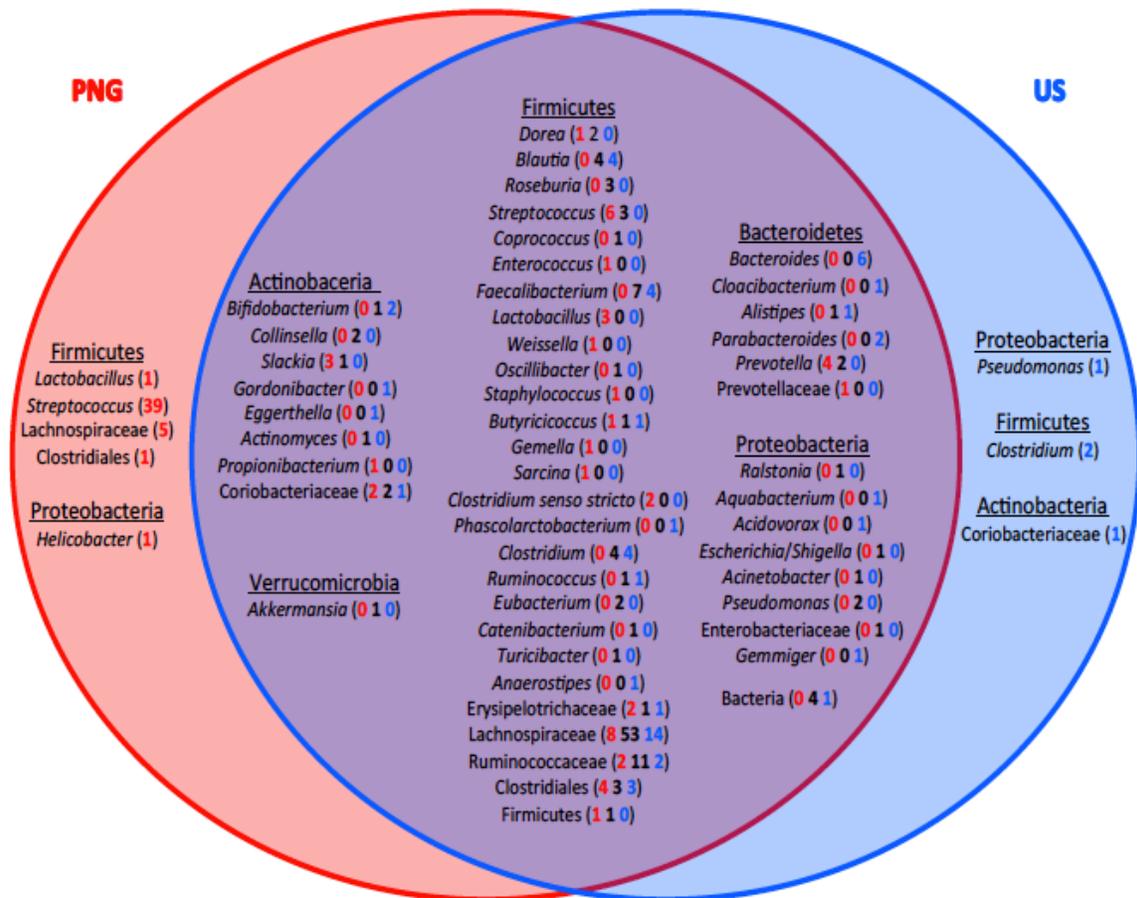
Supplemental Figure S2 - Related to Figure 3. Abundance and phylogenetic diversity of the streptococci OTUs in the PNG and US samples. Bar plots representing the fraction of sequences constituting OTUs taxonomically assigned to the genus *Streptococcus* a proportion of the total number of *Streptococcus* sequences in each sample (A upper panel), or the total number of sequences in each sample (A lower panel). Representative sequences of these OTUs, together with the closest type strains, were aligned with ClustalW and a phylogenetic tree was constructed with neighbor-joining algorithm (MEGA 5.2) (B). Colored squares next to the phylogenetic tree indicate the most abundant streptococci OTUs in the PNG and US fecal samples with matching colors to those of the bar graph of upper panel A. * Indicates OTUs detected in both PNG and US individuals.

Supplemental Figure S3 - Related to Figure 3.



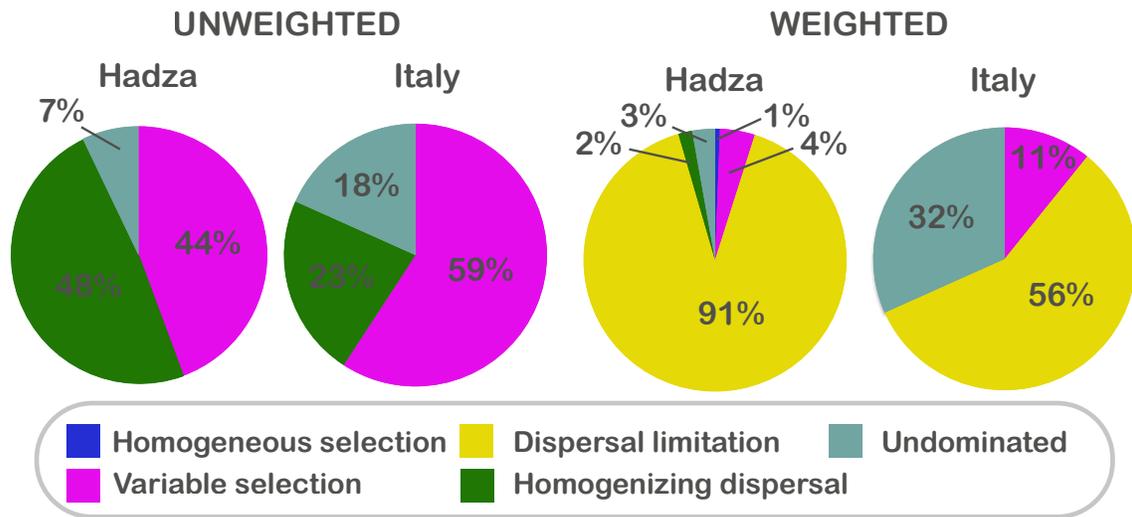
Supplemental Figure S3 - Related to Figure 3. Compositional comparison of the microbiome of Hadza/Italians and Malawians/Venezuela-Amerindians/US. Microbiomes were analyzed to determine OTUs that were shared between the cohorts. The average fraction of shared OTUs (dashed lines) and the average proportion of total sequences that they represent (filled lines) is plotted as a function of the fraction of individuals within that cohort. (A) Hadza/Italy (Schnorr et al., 2014). (B) Malawi/Venezuelan-Amerindian/US (Yatsunencko et al., 2012).

Supplemental Figure S4 - Related to Figure 3.



Supplemental Figure S4 - Related to Figure 3. Venn diagram of shared and exclusive bacterial OTUs in PNG and US fecal samples. OTUs were identified that belonged to the core microbiome in the US and the PNG dataset (present in $\geq 50\%$ of individuals in each cohort). OTUs were taxonomically assigned to a deepest taxonomic level possible, and the number of OTUs for each taxon is shown in brackets. Wilcoxon signed-rank tests and FDR-correction of p-values were conducted to determine OTUs with differing abundance between the PNG and US populations. OTUs are shown in blue if their abundance was significantly higher in the US cohort, red if they were significantly higher in the PNG group, or black if there was no statistical significance in their abundance.

Supplemental Figure S5 - Related to Figure 4.



Supplemental Figure S5 - Related to Figure 4. Summary of the contribution of the ecological processes that determine gut community assembly in Hadza and Italian microbiomes. Pie charts illustrate the fraction that selection dispersal and drift (undominated) contribute to the community assembly process when considering only presence/absence OTU patterns (unweighted) and their abundance (weighted) observed in the dataset.

Supplementary Table S1_ Related to Figure 1. Diet of Asaro and Sausi people based on a dietary survey conducted through interviews with study participants*.

	Asaro	Sausi
Sources of carbohydrate	<ol style="list-style-type: none"> 1. Sweet potato (kaukau) 2. Plantain 3. Cassava (tapiok) 4. Rice 5. Sago 	<ol style="list-style-type: none"> 1. Sweet potato 2. Taro 3. Banana 4. Cassava 5. Yam <p>Others: sago, rice, plantain, pumpkin</p>
Sources of protein**	<ol style="list-style-type: none"> 1. Tinned fish 2. Pork 3. Tinned meat 4. Chicken (karkaruk) 5. Lamb flaps 	<ol style="list-style-type: none"> 1. Pork 2. Fresh fish 3. Chicken 4. Tinned fish 5. Tinned meat <p>Others: cassowary (muruk), bandicoot</p>
Fruits, vegetables and legumes	<ol style="list-style-type: none"> 1. Bush kumu (leafy green) 2. Aibica (type of kumu) 3. Beans 4. Ripe bananas 5. Peanuts <p>Others: pumpkin leaves, guava, pawpaw, tamarillo, corn, cucumber, avocado, karuka nuts</p>	<ol style="list-style-type: none"> 1. Aibica 2. Bush kumu 3. Tulip (leafy green) 4. Beans 5. Pawpaw <p>Others: pineapple, coconut, peanuts, aupa, mango, guava, mon, cucumber, melon, sago leaves, pumpkin leaves, pitpit, cabbage, sugar cane</p>

* Information provided by all 20 Sausi study participants and 8 out of the 20 Asaro ones.

** Proteins are usually consumed 2 times per week.

Supplementary Table S4_ Related to Figure 3. Total, shared, and core (within and across cohorts) OTUs in the fecal microbiota of individuals in PNG, Tanzania (Hadza hunter-gatherers), Malawi, Venezuela (Amerindians), and the Western controls.

Study	Cohort (number of individuals)	Total OTUs detected	Number of OTUs shared between cohorts (proportion of total sequences that they represent mean \pm sd)		OTUs detected in $\geq 50\%$ of subjects within a cohort (proportion of total sequences that they represent mean \pm sd)		Number of OTUs in $\geq 50\%$ of subjects across cohorts (proportion of total sequences that they represent mean \pm sd)	
This study	PNG (40)	1520	664	(87 \pm 5%)	186 (78 \pm 9%)	85	(49 \pm 12%)	
	US (22)			(97 \pm 2%)	169 (86 \pm 12%)		(61 \pm 14%)	
Schnorr et al., 2014	Hadza (27)	1149	398	(80 \pm 5%)	303 (88 \pm 5%)	85	(34 \pm 9%)	
	Italy (16)			(90 \pm 6%)	167 (83 \pm 8%)		(61 \pm 12%)	
	Amerindian (23)			(85 \pm 12%)	1031 (94 \pm 4%)		(49 \pm 13%)	
Yatsunenکو et al., 2012	Malawi (21)	3281	1663	(81 \pm 10%)	1109 (92 \pm 5%)	420	(39 \pm 8%)	
	US (21)			(94 \pm 4%)	770 (87 \pm 10%)		(74 \pm 11%)	

Supplementary Table S5_ Related to Figure 3. Sequence homologues of PNG specific *Lactobacillus* and *Helicobacter* OTUs in fecal microbiomes of other non-industrialized populations.

	Hazda/Italy Schnorr et al., 2014	Malawi/Amerindian/US Yatsunenکو et al., 2012	Burkina Faso/Italy De Fillipo et al., 2010
<i>Lactobacillus</i>	No OTUs were exclusively detected in this cohort	No OTUs were exclusively detected in this cohort	6 Burkina Faso 5 Burkina Faso 1 Burkina Faso + 2 Italy
Closest match (ID %)			<i>L. fermentum</i> KJ54288.1 (100%), <i>L. ruminis</i> CP003032.1 (100%), <i>L. salivarius</i> AB932533.1 (100%)
<i>Helicobacter</i>	2* Hadza	3 Malawian 4 Amerindian	ND
Closest match (ID %)	<i>Helicobacter</i> sp. KC894695.1 (100%)	<i>H. cinaedi</i> KC616311.1 (100%) [†] <i>H. pollorum</i> NR_116595.1 (100%) ^{††}	

* 1 sequence in each individual

† In 2 Malawian and 1 Venezuelan-Amerindian individuals

†† In 1 individual