MICROBIOME

Robust variation in infant gut microbiome assembly across a spectrum of lifestyles

Matthew R. Olm[1]†, Dylan Dahan[1]†, Matthew M. Carter[1], Bryan D. Merrill[1], Feiqiao B. Yu[2], Sunit Jain[2], Xiandong Meng[3], Surya Tripathi[4], Hannah Wastyk[1], Norma Neff[2], Susan Holmes[1,5], Erica D. Sonnenburg[1], Aashish R. Jha[6], Justin L. Sonnenburg[1,2]*

Infant microbiome assembly has been intensely studied in infants from industrialized nations, but little is known about this process in nonindustrialized populations. We deeply sequenced infant stool samples from the Hadza hunter-gatherers of Tanzania and analyzed them in a global meta-analysis. Infant microbiomes develop along lifestyle-associated trajectories, with more than 20% of genomes detected in the Hadza infant gut representing novel species. Industrialized infants—even those who are breastfed—have microbiomes characterized by a paucity of *Bifidobacterium infantis* and gene cassettes involved in human milk utilization. Strains within lifestyle-associated taxonomic groups are shared between mother-infant dyads, consistent with early life inheritance of lifestyle-shaped microbiomes. The population-specific differences in infant microbiome composition and function underscore the importance of studying microbiomes from people outside of wealthy, industrialized nations.

The human gut microbiome undergoes a complex process of assembly beginning immediately after birth (*1*). New microbes attempting to engraft within this community often depend upon niches established by previous colonizing species and thus the final adult microbiome composition may be contingent upon the species acquired early in life. The microbiome assembly process of infants living in industrialized nations is well studied and tends to follow a series of characterized steps that lead to the low-diversity gut microbiome composition characteristic of industrialized adults (*2*). The microbiome assembly process that occurs in infants living nonindustrialized lifestyles (which results in the characteristically diverse adult microbiomes of nonindustrialized adults) (*3*) is largely unknown (*4*). Of particular interest are the following: the timing at which the microbiomes of infants from different lifestyles diverge, the microbes and functions that are characteristic of infants from different lifestyles, and whether there are differences in the taxa that are vertically transmitted from mothers to infants, which seed the microbiome assembly process.

To address these questions we performed metagenomic sequencing on infant fecal samples from the Hadza, a group of modern hunter-gatherers in sub-Saharan Africa (*5*, *6*). The Hadza inhabit seminomadic bush camps of ~5 to 30 people, exhibit a moderate level of community child rearing within these camps (*7*), and are breastfed early in life and weaned onto a diet of baobab powder and premasticated meat at ~2 to 3 years of age (*8*, *9*). In this study we (i) curated and analyzed a global dataset of 1900 16S rRNA sequencing samples of healthy infant fecal samples from 18 populations (including 62 Hadza infant samples) (*2*, *3*, *5*, *10*–*14*) to contextualize the Hadza infant microbiome (figs. S1 and S2), and (ii) performed deep metagenomic sequencing on 39 Hadza infant fecal samples and corresponding maternal fecal samples for 23 infants in order to assess subspecies variation, functional potential, and patterns of vertical transmission (tables S1 and S2).

A UniFrac ordination created from all 1900 16S rRNA sequencing samples revealed age and lifestyle to be strongly associated with the first and second axes of variation, respectively (Fig. 1A) (EnvFit; n = 1900; R^2 = 0.43 and 0.50; P = 0.001 and 0.001). Comparing populations that practice different lifestyles within the same country demonstrates that shared lifestyle affects microbiota composition more than geographic proximity (Fig. 1A, right panel, and fig. S3). The microbiome of infants living industrialized lifestyles diverges from others within the first 6 months of life, whereas the microbiomes of infants living transitional versus nonindustrialized lifestyles diverge at ~30 months of life (Fig. 1B). DNA extraction methods, differences in feeding practices, or other study-specific aspects may contribute to some of the variation in data. Intermediate trajectories are exhibited by populations on the boundaries of industrialized or nonindustrialized lifestyles (Fig. 1B, dashed lines), highlighting the apparent sensitivity of infant microbiota development to lifestyle-related factors.

We identified five microbial coabundance groups (CAGs) (*15*, *16*) in our dataset, which together account for an average of 93.8% of the microbiota composition per sample (Fig. 1C and fig. S4). The *Bifidobacterium-Streptococcus* CAG dominates infants from all lifestyles in early life (0 to 6 months), and over time this CAG yields to the *Bacteroides-Ruminococcus gnavus* CAG in industrialized infants and the *Prevotella-Faecalibacterium* CAG in infants living transitional or nonindustrialized lifestyles (Fig. 1C). Lifestyle-related differences in dominant CAGs become more pronounced over time and mirror taxonomic trade-offs observed in late infancy (*17*) and adulthood (*5*).

We next used our deep metagenomic sequencing data to assess microbiome-encoded functional differences between lifestyles. Broad lifestyle and age associated differences are seen in the overall functional capacity of the infant microbiomes (Fig. 2A), consistent with 16S rRNA amplicon-based analysis (Fig. 1A). Hadza infant metagenomes were assembled and binned into metagenome-assembled genomes (MAGs) representing 745 species, 175 (23.4%) of which represent novel species compared to the Unified Human Gastrointestinal Genome (UHGG) collection (*18*) (table S3). Novel species were recovered from diverse phylogenetic groups (fig. S5A); 88.6% (n = 155) were recovered from multiple Hadza samples (fig. S5B) and their genome quality was observed to be similar to that of genomes in the UHGG (fig. S5C). To assess prevalence through read mapping, MAGs were integrated with genomes recovered from Hadza adults (*19*) and public genomes from the human gut (*18*) into a comprehensive database of 5755 species-representative genomes. Overall, 23.4% of microbial species detected in the Hadza infants represent novel species (table S4). These data support that—similar to the adult Hadza gut—the Hadza infant gut contains extensive previously uncharacterized diversity.

The taxonomic specificity afforded by metagenomic sequencing allowed us to identify particular species that are depleted or enriched in infants living industrialized versus nonindustrialized lifestyles. Identified among the infants in this analysis were 310 VANISH (Volatile and/or Negatively associated in Industrialized Societies of Humans) and 12 BloSSUM (Bloom or Selected in Societies of Urbanization/Modernization) species (table S5 and fig. S6). Comparison against a large database of microbial species from nonhuman habitats (*20*) revealed that no VANISH and only one BloSSUM species match genomes recovered outside of the digestive tract or industrial wastewater, whereas 21 VANISH and three BloSSUM species match microbes recovered from nonhuman animal feces (table S6). VANISH species are more numerous and abundant than BloSSUM (fig. S7), and 63 VANISH species are effectively extinct (never detected) in infants living industrialized or transitional lifestyles. Many VANISH species (45.2%; 140 of 310) are

[1]Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA, USA. [2]Chan Zuckerberg Biohub, San Francisco, CA, USA. [3]ChEM-H Institute, Stanford University, Stanford, CA 94305, USA. [4]Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA, USA. [5]Department of Statistics, Stanford University, Stanford, CA, USA. [6]Genetic Heritage Group, Program in Biology, New York University Abu Dhabi, Abu Dhabi,United Arab Emirates.
*Corresponding author. jsonnenburg@stanford.edu
†These authors contributed equally to this work.

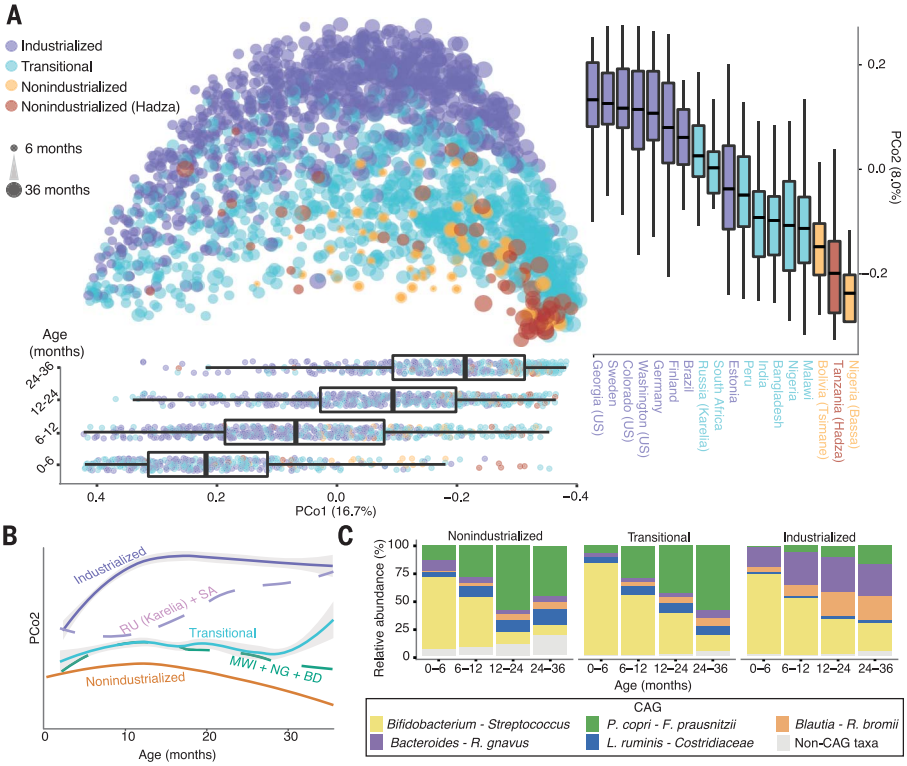

Fig. 1. Age and lifestyle are associated with infant microbiome composition. (**A**) Unweighted UniFrac dissimilarity Principal Coordinates Analysis (PCoA) (top left panel) of 1900 fecal samples from infants (<3 years old) across 18 populations based on amplicon sequence variant abundance. Point color indicates lifestyle and point size is proportional to age in months. Boxplots show the distribution of indicated age groups along PCo1 (bottom) and cohorts along PCo2 (right). (**B**) PCo2 versus sample age for the three lifestyle categories (solid lines) and specific indicated subpopulations (dashed lines). The purple dashed line includes Russia (Karelia) and South Africa [RU (Karelia) + SA] and the green dashed line includes Malawi, Nigeria (Urban), and Bangladesh (MWI + NG + BD). The middle transitional line (blue) contains all transitional samples. Lines are the smoothed conditional mean of PCo2 loadings (loess fit). (**C**) Relative abundance of CAGs by age group and lifestyle. Taxa in annotation are the most abundant taxa in a CAG.

present at 0 to 6 months in nonindustrialized infants whereas BloSSUM species are rarely detected this early in industrialized lifestyle infants (16.7%; 2 of 12) (Fig. 2B). Together these patterns suggest that more species are lost than gained as lifestyles industrialize.

Amplicon and metagenomic data both show that *Bifidobacterium* is the most prevalent taxon in early life (Figs. 1C and 2B). In the first 6 months, infants living nonindustrial lifestyles are dominated by *Bifidobacterium infantis* (also known as *Bifidobacterium longum* subsp. *infantis*) (Fig. 2C), a prolific utilizer of human milk oligosaccharides (HMOs) that is positively associated with human health and commonly used in probiotic supplements (*21*). *B. infantis* is significantly depleted in industrial microbiomes at 0 to 6 months ($P = 0.04$; $n = 151$ industrialized infants; $n = 27$ nonindustrial infants; Wilcoxon rank-sum test) and found at intermediate levels in transitional infants (Fig. 2C). *Bifidobacterium breve*, a species capable of limited HMO degradation (*22*), is instead the most abundant *Bifidobacterium* species in industrialized infants (Fig. 2C). *B. infantis* is antiassociated with *B. breve* in infants across all lifestyles (Fig. 2D). This trend also holds specifically among industrialized infants (correlation = −0.41, $P = 1.0 \times 10^{-3}$, $n =$ 62 industrialized infants, Spearman two sided hypothesis test), suggesting it may be driven by competitive exclusion rather than lifestyle-specific factors.

To determine whether these species-level differences result in community-wide differences in HMO degradation capacity, we mapped our metagenomic reads to the most well-characterized genetic clusters for human milk utilization (table S7). Five of these clusters are involved in HMO degradation (H1 to H5) and one is involved in nitrogen scavenging (referred to as the "urease" cluster) (*21*, *23*); recent studies have linked their expression in the infant gut microbiome to systemic immunological health outcomes (*24*). Five of the six clusters are more prevalent in nonindustrialized than industrialized infants, and their prevalence among transitional infants occurs between these two extremes (Fig. 2E). The H5 cluster, however, exhibits continued persistence beyond the first year of life only in infants from industrialized lifestyles (Fig. 2E). The H5 cluster encodes an ABC-type transporter known to bind core HMO structures, and it is more commonly found in *B. breve* than *B. infantis* (present in 119 of 129 *B. breve* MAGs and 41 of 69 *B. infantis* MAGs recovered from industrialized infants; $P = 1.4 \times 10^{-9}$, Fisher's exact test). The persistence of the H5 cluster beyond 12 months in industrialized infants—a time period in which breastfeeding is less common in these populations—suggests this cassette of genes exists in genomes that are not reliant upon breastfeeding. Breast milk consumption among industrialized infants reduces—but does not eliminate—lifestyle-associated differences in *B. infantis* and HMO-degradation cassette prevalence (fig. S8).

We next investigated strain-level differences among *B. infantis* genomes recovered from infants aged 0 to 1 years old ($n = 96$ MAGs). Several lifestyle-associated functional differences were discovered including (i) enrichment of glycoside hydrolase family 163 (GH163), a CAZyme involved in the utilization of complex N-glycans (including those found on immunoglobulins), in nonindustrialized versus industrialized infants (*25*) (fig. S9, A and B), (ii) differential prevalence of three Pfams (including one related to flagellar assembly) (fig. S9C), and (iii) increased prevalence of four uncharacterized gene clusters in MAGs from nonindustrialized versus industrialized infants (fig. S9D). To verify these metagenomics-based findings, we isolated and sequenced 20 *B. infantis* strains from the same Hadza infant fecal samples (table S3). GH163 and all four gene clusters also showed enrichment among Hadza *B. infantis* isolates as compared to the public reference genomes (fig. S9). Finally, strong lifestyle-specific phylogenetic clustering was observed among *B. infantis* isolate sequences and MAGs (Fig. 2F). This observation of strong region-specific phylogenetic signals could reflect long-term, multigenerational vertical transmission (*26*).

To assess the extent of vertical strain transmission in the Hadza infants, we deeply sequenced fecal samples from corresponding Hadza mothers ($n = 23$ Hadza dyads). Detailed strain-tracking analysis was performed with inStrain (*27*) with a threshold for identical strains of 99.999% popANI (table S8). Dyad pairs share far more strains (6.4 versus 0.3)

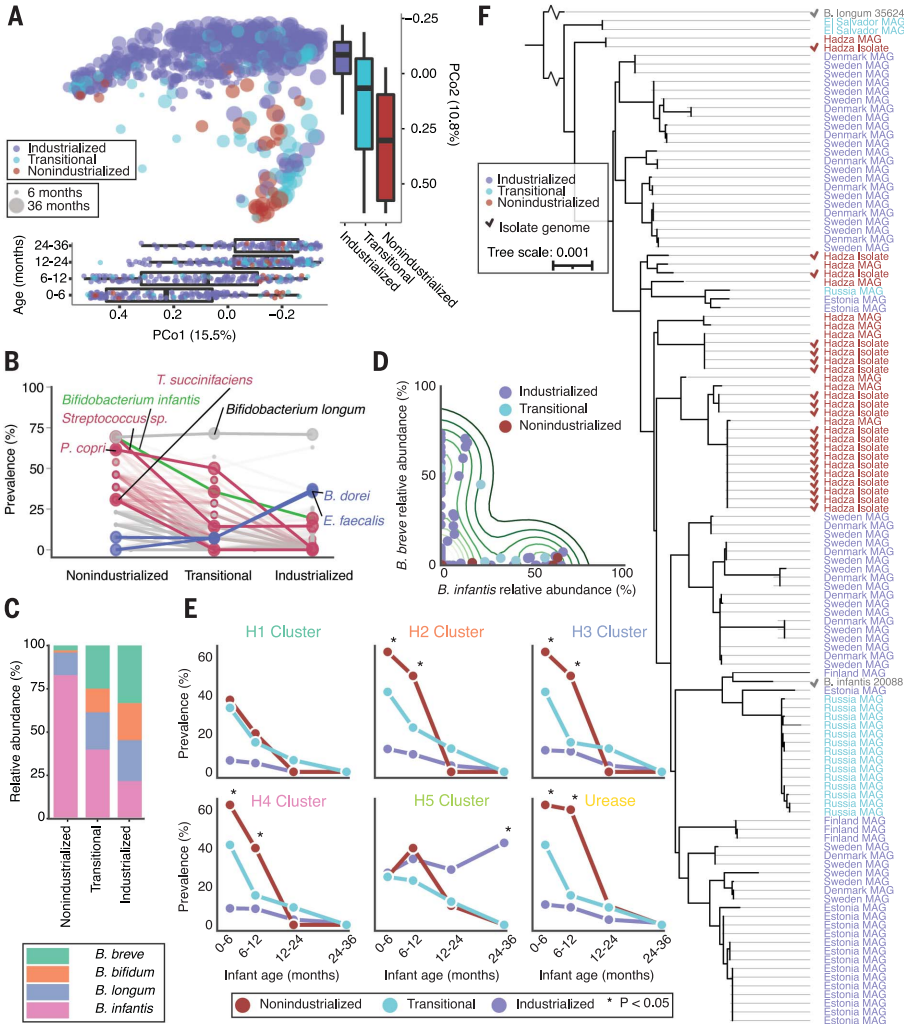

Fig. 2. Age and lifestyle are associated with infant microbiome functions. (**A**) PCoA on the basis of on 682 infant fecal metagenomes described at the gene abundance level in reads per kilobase million (RPKM). Points are colored by lifestyle and point size indicates infant age in months. Boxplots (bottom) show the distribution of indicated age groups in months along PCo1. Boxplots (right) show the distribution of each lifestyle along PCo2. The main axis of variation in this gene-based ordination is significantly associated with age (EnvFit; R^2 = 0.30; n = 679; P = 0.001) and the second axis of variation is significantly associated with lifestyle (EnvFit; R^2 = 0.35; n = 679; P = 0.001). (**B**) Prevalence of species across lifestyles among infants 0 to 6 months old. VANISH (red and green) and BloSSUM (blue) species with the lowest adj-P values have text annotations. *B. infantis* is shown in orange. "Other" taxa (gray) are those that do not significantly differ according to lifestyle. (**C**) Relative representation of four common *Bifidobacterium* species in infants 0 to 6 months old by lifestyle. (**D**) Scatterplot of *B. infantis* versus *B. breve* abundance among infants 0 to 6 months old. Contour lines display the kernel density estimation. (**E**) Prevalence of HMO-utilization clusters across ages and lifestyles. Clusters are considered present if all genes in the cluster are detected above a variable coverage threshold (to ensure that results are robust to differences in sequencing depth; see methods for details). * indicates adj-P < 0.05; Fisher's exact test with false discovery rate correction; nonindustrialized versus industrialized. (**F**) Phylogenetic tree of *B. infantis* genomes based on universal single copy genes. Genome names are colored on the basis of lifestyle of origin. Isolate genomes are marked with a checkmark. Public reference genomes for *B. longum* and *B. infantis* are included (gray text).

and have a higher percentage of strains shared (12.4% versus 0.5%) than nondyad pairs on average (P < 0.01, Wilcoxon rank-sum test) (Fig. 3A). Further, Hadza nondyads living in the same bush camp share more strains than those living in different bush camps (Fig. 3A) (P < 0.01, Wilcoxon rank-sum test), consistent with previously reported increased rates of strain sharing within Fijian social networks (*28*). Vertical strain sharing was detected among a range of phyla in the Hadza (Fig. 3B) and was higher among Bacteroidota and Cyanobacteria and lower among Firmicutes (Fisher's exact test with false discovery rate correction). Industrialized infants also exhibited increased and decreased vertical strain sharing of Bacteroidetes and Firmicutes, respectively (*29*). These results suggest that community interaction during rearing of infants and/or bush camp microenvironments may propagate group microbial sharing (*30*).

The same detailed strain-tracking analysis was next performed on a comparative dataset of 100 dyads from Sweden (*31*). Swedish and Hadza infants were 1.01 ± 0.00 and 0.95 ± 0.21 years old, respectively (P = 0.04, Wilcoxon rank-sum test); in addition, Swedish mothers were sampled immediately after birth whereas Hadza mothers were sampled contemporaneously with infants. Swedish infants born via C-section were excluded from this analysis (n = 17 eliminated) and in silico rarefaction was performed to account for differences in sequencing depth between the studies. Just as *Prevotella* and *Bacteroides* are enriched in nonindustrialized and industrialized infants, respectively (Fig. 1C), *Prevotella* and *Bacteroides* strains are more commonly vertically shared in Hadza and Swedish dyads, respectively (Fig. 3C; Fisher's exact test; P < 0.01). Similar trends are observed for VANISH and BloSSUM taxa (Fig. 3C). The species more abundant in maternal samples were more likely to be vertically transmitted (fig. S10); however, the small difference in infant age between populations may contribute to some differences. The findings suggest that vertical transmission may be a mechanism by which microbiota change is propagated over generations in response to altered lifestyles (*32*–*34*).

Taken together, our data show that infants from all lifestyles begin life with similar *Bifidobacteria*-dominated gut microbiota compositions, but subtle differences detected in early life compound over time. Differences in the species composition and HMO-degradation genes of the initially dominant *Bifidobacterium* communities are especially relevant as recent studies of these same genes suggest that their depletion in industrialized infants could have long-term negative immune consequences (*24*). The same taxa that differentiate lifestyles at 0 to 6 months of life are those that are most commonly vertically transmitted, suggesting that vertical transmission may help establish alternative development trajectories. Crucially, infants living transitional lifestyles display intermediate phenotypes between those of industrialized and nonindustrialized infants in almost all analyses performed. Although not conclusive, this is an important piece of evidence pointing to lifestyle as a possible causative factor in infant microbiome assembly. The Hadza-specific discoveries reported in this work (including the finding of increased nondyad

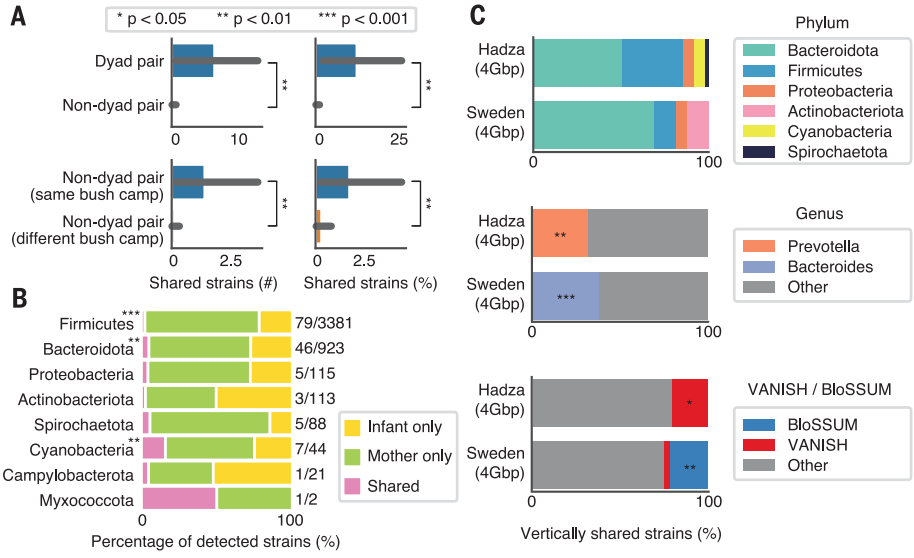

Fig. 3. Strain sharing between mother-infant dyads and nondyads is lifestyle-specific. (**A**) The mean strains shared (left) and the percentage of infant strains found in mothers (right) in mother-infant dyads versus mother-infant nondyads (top) and nondyads from the same bushcamp versus nondyads from different bushcamps (bottom). Error bars represent standard error (*, adj-P < 0.05; **, adj-P < 0.01; ***, adj-P < 0.001; Wilcoxon rank-sum test). (**B**) Percentage of strains detected in all Hadza mothers and infants and whether they are detected in infants only, mothers only, or shared within a mother-infant dyad ("shared") categorized by phylum. Numbers to the right of bars indicate the number of vertically shared strains over the number of strains detected in either infant or maternal samples. Phyla with a significant difference in the percentage of vertically transmitted strains as compared with all other phyla are marked with asterisks (Fisher's exact test with P value correction). (**C**) Percentage of vertically transmitted strains in Hadza and Swedish cohorts by phylum (top), genus (middle; only genera with significant differences shown), and VANISH / BloSSUM (bottom). All metagenomes were subset to 4Gbp for this analysis to remove any biases associated with sequencing depth. Taxa that are significantly enriched in either cohort are marked with an asterisk (*, adj-P < 0.05; **, adj-P < 0.01; ***, adj-P < 0.001; Fisher's exact test).

vertical transmission among members of the same bush camp, a social structure with no equivalent among industrialized communities) exemplify the importance of studying people outside of industrialized nations and highlights the need for additional studies to provide equity in understanding microbiomes across global societies. Our results also highlight the question of whether lifestyle-specific differences in the gut microbiome's developmental trajectory predispose populations to diseases common in the industrialized world, such as those driven by chronic inflammation (*35*, *36*).

REFERENCES AND NOTES

1. T. Yatsunenko *et al.*, *Nature* **486**, 222–227 (2012).
2. C. J. Stewart *et al.*, *Nature* **562**, 583–588 (2018).
3. G. K. Fragiadakis *et al.*, *Gut Microbes* **10**, 216–227 (2019).
4. M. C. de Goffau *et al.*, *Nat. Microbiol.* **7**, 132–144 (2022).
5. S. A. Smits *et al.*, *Science* **357**, 802–806 (2017).
6. F. Marlowe, *The Hadza: Hunter-Gatherers of Tanzania* (Univ. of California Press, 2010).
7. N. G. Blurton Jones *et al.*, *Am. J. Phys. Anthropol.* **89**, 159–181 (1992).
8. A. N. Crittenden, N. L. Conklin-Brittain, D. A. Zes, M. J. Schoeninger, F. W. Marlowe, *Evol. Hum. Behav.* **34**, 299–304 (2013).
9. A. N. Crittenden, F. W. Marlowe, *Hum. Nat.* **19**, 249–262 (2008).
10. A. S. Raman *et al.*, *Science* **365**, eaau4735 (2019).
11. T. Vatanen *et al.*, *Cell* **165**, 842–853 (2016).
12. F. A. Ayeni *et al.*, *Cell Rep.* **23**, 3056–3067 (2018).
13. A. W. Kamng'ona *et al.*, *Sci. Rep.* **9**, 12893 (2019).
14. D. D. Sprockett *et al.*, *Nat. Commun.* **11**, 3772 (2020).
15. S. C. Watts, S. C. Ritchie, M. Inouye, K. E. Holt, *Bioinformatics* **35**, 1064–1066 (2019).
16. A. R. Jha *et al.*, *PLOS Biol.* **16**, e2005396 (2018).
17. J. Roswall *et al.*, *Cell Host Microbe* **29**, 765–776.e3 (2021).
18. A. Almeida *et al.*, *Nat. Biotechnol.* **39**, 105–114 (2021).
19. B. D. Merrill *et al.*, *bioRxiv* (2022), p. 2022.03.30.486478.
20. S. Nayfach *et al.*, *Nat. Biotechnol.* **39**, 499–509 (2021).
21. R. M. Duar *et al.*, *Nutrients* **12**, 3247 (2020).
22. M. Sakanaka *et al.*, *Nutrients* **12**, 71 (2019).
23. R. G. LoCascio, P. Desai, D. A. Sela, B. Weimer, D. A. Mills, *Appl. Environ. Microbiol.* **76**, 7373–7381 (2010).
24. B. M. Henrick *et al.*, *Cell* **184**, 3884–3898.e11 (2021).
25. J. Briliūtė *et al.*, *Nat. Microbiol.* **4**, 1571–1581 (2019).
26. P. Ferretti *et al.*, *Cell Host Microbe* **24**, 133–145.e5 (2018).
27. M. R. Olm *et al.*, *Nat. Biotechnol.* **39**, 727–736 (2021).
28. I. L. Brito *et al.*, *Nat. Microbiol.* **4**, 964–971 (2019).
29. Y. C. Lou *et al.*, *Cell Rep. Med.* **2**, 100393 (2021).
30. A. H. Moeller *et al.*, *Science* **353**, 380–382 (2016).
31. F. Bäckhed *et al.*, *Cell Host Microbe* **17**, 690–703 (2015).
32. E. D. Sonnenburg *et al.*, *Nature* **529**, 212–215 (2016).
33. P. Vangay *et al.*, *Cell* **175**, 962–972.e10 (2018).
34. M. J. Blaser, *Cell* **172**, 1173–1177 (2018).
35. J. L. Sonnenburg, E. D. Sonnenburg, *Science* **366**, eaaw9255 (2019).
36. E. D. Sonnenburg, J. L. Sonnenburg, *Nat. Rev. Microbiol.* **17**, 383–390 (2019).

ACKNOWLEDGMENTS

We acknowledge the numerous people and organizations who provided logistical support and conducted sample collection in the USA, Tanzania, and Nepal, including Dorobo Safaris, the Human Food Project, J. Changalucha, A. Manjurano, M.G. Domiguez-Bello, M. St. Onge, A. Weakly, and Y. Gautam. We thank D. Relman and C. Damman for helpful discussion and input throughout project conceptualization and analysis. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This research utilizes data obtained by the TEDDY study group, a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), the National Institute of Allergy and Infectious Diseases (NIAID), the National Institute of Child Health and Human Development (NICHD), the National Institute of Environmental Health Sciences (NIEHS), Centers for Disease Control and Prevention (CDC), and JDRF. The data from the TEDDY study reported here were supplied by the database of Genotypes and Phenotypes (dbGaP; Study Accession: phs001443.v1.p1), which is maintained by the National Center for Biotechnology Information (NCBI). This manuscript was not prepared in collaboration with investigators of the TEDDY study and does not necessarily reflect the opinions or views of the TEDDY study, dbGaP, or the NIDDK. J.L.S. is a Chan-Zuckerberg Biohub Investigator. **Recognition of work on Indigenous communities:** Research involving Indigenous communities is needed for a variety of reasons, including assurance that scientific discoveries and understanding appropriately represent all populations and do not only benefit those living in industrialized nations. Special consideration must be made to ensure that this research is conducted ethically and in a nonexploitative manner. In this study we performed deep metagenomic sequencing on fecal samples collected from Hadza hunter-gatherers in 2013 and 2014; these samples were analyzed in previous publications with different methods (*3*, *5*). A material transfer agreement with the National Institute for Medical Research in Tanzania ensures that the collected stool samples are used solely for academic purposes, and permission for the study was obtained from the National Institute of Medical Research (MR/53i 100/83, NIMR/HQ/R.8a/Vol.IX/1542) and the Tanzania Commission for Science and Technology. Verbal consent was obtained from the Hadza after the study's intent and scope was described with the help of a translator. The publications that first described these samples included several scientists and Tanzanian field guides as coauthors for the critical roles they played in sample collection; however, no new samples were collected in this study and as such, only scientists who contributed to the analyses described here are included as coauthors in this publication. It is currently not possible for us to travel to Tanzania and present our results to the Hadza people; however, we intend to do so once the conditions of the COVID-19 pandemic allow it. **Funding:** This work was supported by the following: National Institutes of Health grants DP1-AT009892 and R01-DK085025 (to J.L.S.); NSF Graduate Research Fellowship grants DGE-1656518 (to D.D.) and DGE-114747 (to B.D.M.); Stanford Graduate Smith Fellowship (to D.D. and M.M.C.); National Institutes of Health grant F32DK128865 (to M.R.O.); Funding was also provided by the Bill and Melinda Gates Foundation. **Author contributions:** Conceptualization: D.D., A.R.J., and J.L.S. Genomic Sequencing: N.N., B.Y., B.D.M., S.T., and D.D. Methodology: D.D., A.R.J., M.R.O., M.M.C., B.D.M., S.J., S.H., and H.W. Data analysis: D.D., M.R.O., M.M.C., B.D.M., S.T., and S.J. Funding Acquisition: D.D., J.L.S., E.D.S., A.R.J., and M.R.O. Supervision: E.D.S., J.L.S., A.R.J., and S.H. Writing - original draft: D.D., M.R.O., E.D.S., and J.L.S. Writing - reviewing and editing: M.R.O., D.D., A.R.J., E.D.S., J.L.S., and M.M.C. **Competing interests:** Authors declare that they have no competing interests. **Data and materials availability:** The authors declare that the data supporting the findings of this study are available within the paper and its supplementary information files. Metagenomic reads and genomes generated in this study are available under BioProject PRJEB49206. Accession numbers for individual samples and genomes are available in tables S2 and S3.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj2972
Materials and Methods
Figs. S1 to S10
Tables S1 to S8
References (*37*–*71*)
MDAR Reproducibility Checklist
Data S1 to S8

View/request a protocol for this paper from *Bio-protocol*.

Submitted 3 May 2021; resubmitted 27 January 2022
Accepted 5 May 2022
10.1126/science.abj2972