Supplemental Information Guide

Supplemental Table 1. List of Human Metagenomic Project (HMP) metagenomic samples used in this study.
SRA numbers and personal identifiers for all HMP samples used in the study.

Supplemental Table 2. List of FijiCOMP metagenomic samples used in this study.
Sample identifiers for the stool and saliva samples are provided as are their village associations. Village 1 is located in the Macuata province, whereas villages 2-4 are located in the Bua province. Samples from which single-cells were isolated are noted.

Supplemental Table 3. Human Microbiome Project reference genomes used in this study.
The following 387 genomes are part of the HMP Reference Genome set (Human Microbiome Jumpstart Reference Strains Consortium, 2010) and met our criteria for inclusion in the study.

Supplemental Table 4. Single-cell assemblies from the FijiCOMP study used in this analysis.
For each MDA-amplified cell, the following data is listed: identification number; person from which the cell was isolated; taxonomy assigned by RDP and core AMPHORA genes blasted to the NR database; isolation method; read count (post filtering); assembly size; N50; the number of contigs; the percent of initial reads filtered due to human contamination and bacterial contamination; the number of core genes identified using AMPHORA2 and the number whose BLASTp hits were consistent with the assigned taxonomies; the percent completeness and contamination as determined by CheckM; whether the 16S rRNA sequence is available and whether it is full length or just the V68 region; and the number of contigs that were verified by BLASTn with closely related reference genomes. Due to the similarity between *Salmonella* and *Escherichia* genomes, we anticipate that reads deemed "contaminants" in cell #21_2I were likely genomic. Note, metagenomics sequencing was not performed on the donor W2.22.

Supplemental Table 5. Horizontally transferred genes observed in this study.
This spreadsheet comprises all 37,853 horizontally transferred genes identified in the study. 10,461 unique genes remained after redundant genes were removed from the dataset. For each gene, the following metadata is included: whether it is a redundant gene (i.e. it was not used for read alignment) and which gene it is identical or near-identical to; the genome in which it was observed; the gene length; whether it was observed on a contig whose phylogenetic origin could be verified through BLASTn searches within closely related organisms; whether it is situated next to one or more tRNA genes; the maximum number of genomic contexts observed between that gene and other mobile genes in the dataset; a consensus functional annotations; whether it is a carbohydrate metabolizing enzyme (CAZYme); whether it confers antibiotic resistance; whether it is a horizontal transfer machinery gene and, if so, what type; whether vector contamination was observed to overlap with the gene; whether it was included in the higher confidence subset; and its prevalence, median and mean abundance in the HMP and the FijiCOMP cohorts.

Supplemental Table 6. Prevalence and abundance of carbohydrate metabolism enzymes across the HMP and FijiCOMP cohorts.
This table was populated based on the FPKM abundances of each CAZYme gene family across individuals in either the HMP or FijiCOMP. Median values represent the median of individuals harboring those genes. q-values of prevalence comparisons are based on FDR-corrected Fisher's exact tests and for abundance comparisons are based on FDR-corrected Mann-Whitney tests.

Supplemental Table 7. Prevalence and abundance of antibiotic resistance genes, by class, across the HMP and FijiCOMP cohorts.
This table was populated based on the FPKM abundances of each gene family across individuals in either the HMP or FijiCOMP. Median values represent the median of individuals harboring those genes. q-values are based on FDR-corrected Mann-Whitney tests.

Supplemental Table 8. Mobile genes enriched or depleted in a single village.
For each village, a Fisher's exact test was performed to determine whether any gene was associated with any specific village. Mann-Whitney tests were performed to determine whether any specific organism was enriched or depleted in a specific village. q-values represent FDR-adjusted p-values.

Supplemental Table 9. Organisms enriched or depleted in a single village.
For each village, a Mann-Whitney test were performed to determine if any specific organism was enriched or depleted in any specific village. q-values represent FDR-adjusted p-values. Only Village 1 was enriched in organisms. None of the organisms' prevalence was associated with a specific village.

Supplemental Table 10. Horizontally transferred genes adjacent to multiple tRNAs.
Information on each read-pair that was observed as linking a horizontally transferred gene to multiple different adjacent tRNA genes. There were 838 genes that were adjacent to tRNAs. Of these, 194 were found in more than one context (more than one isoacceptor, tRNA gene or bacterial genera). Taxonomies were obtained by searching NT using BLASTn.

Supplemental Text. DNA sequences of all horizontally transferred genes observed in this study.
A FASTA file containing 37,853 DNA sequences. Sequence identifiers correspond to the cell identifier followed by the contig number and the gene number on that contig.