

Legends for Supplementary Tables and Figures

Table S1. Metadata for study

Summary of metadata for all 64 individuals included in this study. For 5 individuals, we collected two fecal samples (7 days apart for 4 samples and 1 day apart for 1 sample). For parasites detection, we identified *Ascaris* (ASC), *Ancylostoma* (ANC), *Trichuris* (TRI) and *Entamoeba* (ENT). When none of these were detected, the individuals were annotated as negative (Neg). The diet questionnaire consists in the assessment of the frequency with which each dietary component is consumed per month (0: never, 1: once a month or less, 2: one to three times a month, 6: once to twice a week, 15: three to five times a week, 30: almost every day). Throughout, NA refers to “Not available.”

Table S2. Significance of covariates describing composition of gut microbiota

Summary of p-values from a PERMANOVA analysis of microbiome composition (OTUs): (i) merged (OTUs collapsed based on shared taxonomic identifier) and filtered ($\geq 0.1\%$ in ≥ 4 individuals), (ii) merged and unfiltered, (iii) partially merged (OTUs collapsed only when family, genus, or species names are shared) and unfiltered, (iii) unmerged and unfiltered, and both weighted and unweighted UniFrac distance matrices for different covariates.

S3 Table. Summary of significant taxa across covariates

Summary of the number of significant taxa ($q < 0.05$; ANOVA) for covariates that were identified as being significant for at least one taxon. Analysis is based on abundances of OTUs that were merged based on taxa names such that abundances for all OTUs with identical identifiers were combined. Merged OTUs = all OTUs with identical identifiers were combined. Partially merged OTUs = only OTUs with identical identifiers at the family, genus, or species levels were combined. For the covariate “locations”, the Ndtoua village where only two individuals were sampled was not considered.

Table S4. Significantly different taxa across *Entamoeba* status

Summary of means and q-values (ANOVA) for all taxa (occurring at $\geq 0.1\%$ in at least 4 individuals) that differ significantly between *Entamoeba* negative (*Ent-*) and positive (*Ent+*) individuals based on merged (OTUs collapsed based on taxa name) and partially merged (OTUs collapsed only when family, genus, or species names are shared) OTU tables. Taxa listed at lower resolution exclude those taxa that are listed at higher resolution (e.g. if Clostridiaceae and Clostridiaceae Clostridium are both listed, abundances of Clostridiaceae.exclude Clostridium). Taxa highlighted in pink are common both merged and unmerged analyses.

Table S5. Significantly different taxa across subsistence groups

Summary of means and q-values (ANOVA) for all taxa ($\geq 0.1\%$ in at least 4 individuals) that differ significantly between subsistence groups based on merged (OTUs collapsed based on taxa name) and partially merged (OTUs collapsed based on taxa name restricted to below the family level) OTU tables. Taxa listed at lower resolution exclude those taxa listed at higher resolution. (e.g. if Clostridiaceae and Clostridiaceae Clostridium are both listed, abundances of Clostridiaceae.exclude Clostridium). Taxa highlighted in pink are common both merged and unmerged analyses. Significant q-values (<0.05) for pairwise comparisons between subsistence groups are highlighted in yellow, and the average number of significant pairwise comparisons is reported for each pair of subsistence groups (Fis = Fishing population; Far(S) = Farmers from the South; Far(N) = Farmers from the North; HG = Hunter-gatherers).

Table S6. Significantly different KEGG pathways across subsistence and *Entamoeba* status

Summary of means and q-values (ANOVA) for all abundant level 3 KEGG pathways ($\geq 0.4\%$ in at least 4 individuals) across *Entamoeba* infection status and subsistence group (Fis: Fishing population; Far(S): Farmers from the South; Far(N): Farmers from the North; HG: Hunter-gatherers). Pathways are ranked by q-values for the infection status.

Table S7. RFC predictions for subsistence

Summary of random forest classifier (RFC) model predictions for four subsistence groups (Fis = fishing population, Far(S) = farmers from the South, Far(N) = farmers from the North, and HG = hunter-gatherers) based on individual gut microbiome composition. A RFC with 2000 decision trees was trained on the data with 5-fold cross-validation.

Figure S1. Parasite prevalence across subsistence groups

Prevalence of individual infection by four parasites (*Trichuris*, *Ascaris*, *Ancylostoma*, and *Entamoeba*) across the four subsistence groups. Fis = fishing population, Far(S) = farmers from the South, Far(N) = farmers from the North, and HG = hunter-gatherers. The total number of samples in each population (n) is indicated below the bar. Parasitism was assessed as either presence or absence of the species surveyed.

Figure S2. Alpha rarefaction curves for across subsistence groups

Alpha rarefaction curves for the four subsistence groups using the Phylogenetic Distance Whole Tree metric for alpha diversity (means across 10 iterations). Fis = fishing population, Far(S) = farmers from the South, Far(N) = farmers from the North, and HG = hunter-gatherers. Variance is among individuals. All further analyses are based on 50,000 reads per sample, the maximal depth with which we do not exclude any individuals.

Figure S3. Beta diversity estimates within individuals, subsistence group, and location

(a) Beta diversity estimates between all pairs of individuals, for pairs within subsistence groups, within locations, and within individuals based on weighted (left panel) and unweighted (right panel) UniFrac distances. *P*-values are from a Wilcoxon Rank Sum test. (b) Multidimensional Scaling (MDS) showing separation among replicate samples for 5 individuals using weighted (left panel) and unweighted (right panel) UniFrac distances. The first two principal components are shown on the axes.

Figure S4. Multidimensional Scaling plots for weighted UniFrac distances for *Entamoeba* status and subsistence

Multidimensional Scaling plots for (a) weighted UniFrac distances colored by *Entamoeba* status (b) weighted UniFrac distances colored by subsistence (Fis = fishing population, Far(S) = farmers from the South, Far(N) = farmers from the North, and HG = hunter-gatherers), and (c) unweighted UniFrac distances colored by subsistence. For each plot, the first two principal components (PC1 and PC2) are shown.

Figure S5. Summary of RFC top ten taxa for *Entamoeba* status

Summary of the ten most discriminating taxa identified by a random forest classifier (RFC) model to be predictive of *Entamoeba* infection status. A RFC with 2000 decision trees was trained on the data (relative abundances of all taxa occurring at $\geq 0.1\%$ in at least 4 individuals) with 5-fold cross-validation. Importance values were calculated as the mean decrease in node impurity. 95% confidence intervals from 1000 random forests are shown. Mean accuracy over the 5 folds was 0.79 (SD = 0.09). $P < 0.001$, estimated using 1000 permutation tests. Taxa that were identified as being significant in an ANOVA are indicated by asterisk(s) (*); $q < 0.0001$ (***), $q < 0.0001$ (**), $q < 0.001$ (*).

Figure S6. Alpha diversity across *Entamoeba* status

Comparison of alpha diversity for *Entamoeba* negative (*Ent-*) and positive (*Ent+*) individuals using the Shannon Index (a) and Simpson's Index (b). P-values are based on a Welch's t-test.

Figure S7. Beta diversity across *Entamoeba* status

Comparison of beta diversity for *Entamoeba* negative (*Ent-*), positive (*Ent+*), and between *Ent-* and *Ent+* individuals based on weighted UniFrac distances. P-values are based on a Welch's t-test.

Figure S8. Phyla significantly correlated with alpha diversity across *Entamoeba* status

Phyla (occurring at $\geq 0.1\%$ in at least 4 individuals) for which abundance is significantly correlated with alpha diversity (phylogenetic distance whole tree metric) in *Entamoeba* negative (*Ent-*, blue) and positive (*Ent+*, red) individuals. $q < 0.05$; ANOVA.

Figure S9. Summary of predicted metagenomic composition across *Entamoeba* status

Summary of predicted metagenomic composition for *Entamoeba* negative (*Ent-*) and positive (*Ent+*) individuals based on phylotypic investigation of communities by reconstruction of unobserved states (PICRUSt). Relative abundance of (a) the most abundant ($\geq 0.4\%$ in at least one group) and (b) least abundant ($\leq 0.1\%$ in at least one group) of KEGG (Level 3) pathways.

Figure S10. Age of *Entamoeba* positive and negative individuals

Mean age of *Entamoeba* negative (*Ent-*) and positive (*Ent+*) individuals. P-value are based on a Welch's t-test.

Figure S11. Summary of relative abundance of taxa across *Entamoeba* status

Summary of the relative abundance of taxa ($\geq 0.1\%$ in at least 4 individuals) for the number of infecting parasites in the absence of *Entamoeba* infection (0-3). Taxa are colored by phylum (Actinobacteria (Act.) = red, Bacteroidetes (Bact.) = green, Cyanobacteria (Cyan.) = black, Elusimicrobia (Elus.) = gold, Firmicutes (Firm.) = blue, Fusobacteria (Fus.) = pink, Lentisphaerae (Lent.) = yellow, Proteobacteria (Prot.) = purple, Spirochaetes (Spir.) = orange, and Tenericutes (Ten.) = gray). The number of individuals (N) that falls into each category is indicated.

Figure S12. Alpha diversity across infecting parasite richness

Alpha diversity across individuals grouped by the number of infecting parasite species in the absence of *Entamoeba* infection (0-3) using the phylogenetic distance whole tree metric. P-values are based on a Welch's t-test.

Figure S13. Summary of relative abundance of taxa across subsistence

Summary of the relative abundance of taxa (occurring $\geq 0.1\%$ in at least 4 individuals) for individuals across locations. The Ndtoua village is not considered here as it only includes two individuals. Taxa are colored by phylum (Actinobacteria (Act.) = red, Bacteroidetes (Bact.) = green, Cyanobacteria (Cyan.) = black, Firmicutes (Firm) = blue, Fusobacteria (Fus.) = pink, Lentisphaerae (Lent.) = yellow, Proteobacteria (Prot.) = purple, Spirochaetes (Spir.) = orange, and Tenericutes (Ten.) = gray). Ebo = Ebodie; Aka = Akak; Afa = Afan Essokie; Ban = Bandevouri; Mak = Makoure; Bid = Bidou. Fis = Fishing population; Far(S) = Farmers from the South; Far(N) = Farmers from the North; HG = Hunter-gatherers.

Figure S14. Random forest classifier predictions for subsistence

Summary of the ten most discriminating taxa identified by a random forest classifier (RFC) model to be predictive of subsistence group. A RFC with 2000 decision trees was trained on the

data (relative abundances of all taxa occurring at $\geq 0.1\%$ in at least 4 individuals) with 5-fold cross-validation. (a) Bar plot summarizing the predictions of the model for each population. Fis = Fishing population; Far(S) = Farmers from the South; Far(N) = Farmers from the North; HG = Hunter-gatherers. The percent accuracy for each population is indicated above the bars. (b) The ten most important taxa identified in the model ranked by their median importance value. Importance values were calculated as the mean decrease in node impurity. 95% confidence intervals from 1000 random forests are shown. Mean accuracy over the 5 folds was 0.59. $P < 0.001$, estimated using 100 permutation tests. Taxa that were identified as being significant in an ANOVA are indicated by asterisk(s) (*); $q < 0.0001$ (***), $q < 0.0001$ (**), $q < 0.001$ (*). (c) Normalized relative abundances across subsistence groups of the four taxa (amongst the top ten identified in the RFC model) that are members of the Lachnospiraceae family. q-values are based on an ANOVA.

Figure S15. Shannon and Simpson's diversity across subsistence

Comparison of individual gut microbial diversity across subsistence modes based on (a) the Shannon Index and (b) the Simpson's Index metrics for alpha diversity. P-values are based on Welch's t-tests. Fis = Fishing population; Far(S) = Farmers from the South; Far(N) = Farmers from the North; HG = Hunter-gatherers.

Figure S16. Phylogenetic, Shannon, and Simpson's diversity across location

Comparison of individual gut microbial diversity across locations based on (a) the Phylogenetic Distance Whole Tree metric, (b) the Simpson's Index, and (c) the Simpson's Index. P-values are based on Welch's t-tests. There are only 2 individuals from Ndtoua (hunter-gatherers) so they were excluded from analyses by location. Ebo = Ebodie; Aka = Akak; Afa = Afan Essokie; Ban = Bandevouri; Mak = Makoure; Bid = Bidou. Fis = Fishing population; Far(S) = Farmers from the South; Far(N) = Farmers from the North; HG = Hunter-gatherers.

Figure S17. Weighted UniFrac distance across subsistence and location

Beta diversity between individuals based on weighted UniFrac distances across (a) subsistence modes and (b) location. (c) Beta diversity between pairs of subsistence groups. All p-values are based on Welch's t-tests. Subsistence: Fis = Fishing population (blue); Far(S) = Farmers from the South (light green); Far(N) = Farmers from the North (dark green); HG = Hunter-gatherers (red). Locations: Ebo = Ebodie; Aka = Akak; Afa = Afan Essokie; Ban = Bandevouri; Mak = Makoure; Bid = Bidou. There are only 2 individuals from Ndtoua (hunter-gatherers) so they were excluded from analyses by location, but were included in those for subsistence.

Figure S18. Predicted metagenomic composition of gut microbiota across subsistence mode

Summary of predicted metagenomic composition for individuals across subsistence modes based on phylotypic investigation of communities by reconstruction of unobserved states (PICRUSt). Relative abundance of (a) the most abundant (>0.4% in at least one group) and (b) least abundant (<0.1% in at least one group) of KEGG (Level 3) pathways. Fis = Fishing population; Far(S) = Farmers from the South; Far(N) = Farmers from the North; HG = Hunter-gatherers.