**Sample Collection**

*Village identification*

*Sample Stratification*

Human fecal samples (n = 120) were collected from children and adults ranging in age from 5 to 87 years old. Sample collection was stratified by age, sex, and multifamily quartier (compound). A total of 11 quartiers were sampled, and within each quartier a range of 1 to 4 families were sampled, resulting in a total of 30 families sampled. Four samples were collected from each family: a child (male or female), an elder male, an adult male, and adult female. This resulted in a total of 120 human fecal samples.

Pig samples were collected from each family that donated human samples. Samples came from an adult male, adult female, and juvenile of either sex, resulting in a total of 90 pig samples.

*Sample Collection and Storage*

In the field, stool samples were in aliquoted into 30 mL storage tubes and placed in an ice chest, before being moved to a -20 °C freezer each night. After all samples were collected, samples were transported on ice to Centre MURAZ and stored at -80 °C until extraction.

*Human Survey*

A 200-question metadata survey was completed by each individual about lifestyle practices. Questions covered information about place of birth, age, amount of schooling, water source/treatment, number of livestock, dietary recall, source of food, medication recall, and activity (see table N).

*Pig Survey*

**Laboratory Methods**

*Initial Extraction*

Human and pig fecal samples were extracted at the Centre Muraz Research Institute in Burkina Faso with Qiagen’s AllPrep PowerViral DNA/RNA isolation kit, following the manufacturer’s guidelines. Extracts and fecal samples were stored at -80 °C and then shipped frozen to the Laboratories of Molecular Anthropology and Microbiome Research (LMAMR) at the University of Oklahoma. The extract from individual 24\_04 broke during shipment and thus we do not have data from this extraction at Centre MURAZ. The human fecal samples were processed anaerobically to provide future opportunity to culture anaerobic bacteria. In brief, samples were weighed upon arrival and 2 grams of fecal sample was added to 10 mL of anoxic Phosphate Buffered Saline (PBS) to create a fecal slurry. Slurry aliquots were added to glycerol stocks in serum bottles for anaerobic culturing, while separate aliquots of the slurry were set-aside to be used for Propidium Monoazide (PMA) Live/Dead screening. Additionally, about 0.5 g of raw fecal sample was aliquoted to allow for re-extraction of the raw sample – as the raw fecal sample is what was extracted at Centre Muraz. All aliquots and the remaining raw sample were stored in -80 °C until use. See Table N for mass of each sample upon arrival at LMAMR and mass used for aliquots.

*Re-extraction*

Twenty samples (table N) were chosen to be extracted from the raw fecal aliquots at LMAMR to determine effect of transport and storage on the fecal microbiome community. The samples were selected from the subset of samples extracted by DJ at Centre MURAZ (multiple individuals extracted the samples at Centre MURAZ) to best control for user variability in extraction. The Qiagen AllPrep PowerViral DNA/RNA isolation kit was used for extraction.

*PMA treatment*

The same twenty samples used for re-extraction were selected for PMA treatment and extraction. From an aliquot of the fecal slurry, 250ul was extracted without PMA treatment while a separate 250ul of fecal slurry was used for PMA treatment. PMA was diluted to 10mM, and 2.5ul was added to the 250ul fecal slurry and vortexed to mix. Samples were incubated at room temperature for 15 minutes and then placed in a blue light box to activate the PMA for 15 minutes. After PMA treatment, samples were extracted with Qiagen’s AllPrep PowerViral DNA/RNA isolation kit.

Samples were named in the following manner: population.FamilyNumberIndividualNumber.Treatment. All Burkina Faso samples come from the same population and thus share “bftm” at the start of the sample name. The family number is a two-digit number referring to the family the individual comes from, ordered sequentially (e.g those from the first family sample are 01, those from the second family are 02, etc.). The Individual number refers to the individual within the family (the first individual from each family is 01, the second from each family is 02, etc.). Treatment refers to either the location of extraction or whether the sample was treated with PMA. Samples extracted at Centre Muraz are labelled “MURAZ”, those re-extracted from raw feces are labelled “LMAMR”, the samples with PMA treatment are labelled “wPMA” and the matching samples extracted from the slurry without PMA are labelled “noPMA”.

The V4 region of the 16S rRNA gene was quantified using qPCR and samples were diluted to the same concentration. In short, a serial dilution *Escherichia* coli standards and each extract were amplified with primers. Quantification was performed using the Roche LightCycler software, see tableN for sample concentration. Samples were grouped by similar concentration (no greater than a 10 fold difference in concentration), diluted to the same concentration and PCR amplified in triplicate with a unique Golay barcode for each sample. PCR products were pooled in equimolar concentrations, purified with Minelute spin column, and sequenced on an Illumina MiSeq with a 2x250 kit.

**Bioinformatic processing**

Sequence reads were demultiplexed with bcl2fastq. AdapterRemoval2 was used to trim low quality reads and merge reads from the same sample. USEARCH was used to eliminate singleton sequences then identify and cluster true biological sequences (zOTUs) as well as traditional OTUs (at 97% similarity).

For the poster abstract

Custom scripts (Python and R) were used to apply a non-parametric ANOVA (Kruskall-Wallis) test to compare taxon abundance between metadata groups at the 97% OTU level. Expected associations between diet and taxon abundance were found, such as individuals with high fiber diets demonstrating an increased abundance *Prevotella­,* a known fiber-digestor. Otherwise, the only significant associations found (fdr adjusted P-value < 0.05) were between various taxa and quartier. There were similar associations with family but it was with fewer taxa and the associations were not as strong. The observed trend was that *Prevotella* was inversely related to *Oscillibacter, Collinsella, Sporobacter*, and *Subdoligranulum.* Existing literature indicates that the taxa inversely associated with *Prevotella* are found at increased abundance in those Type 2 diabetes and high fat diets in mice (need lit). Looking at quartiers, there does seem to be higher proportion of individuals from the high *Prevotella* quartiers that eat high fiber foods, compared to those from the quartiers with high prevalence of the other taxa.

Compar.r script to find associations between metadata groups and taxon abundance, quartier had the most and strongest associations (family had similar taxa but less strong associations). Green seeds in 24hrs had prevotella association but no other metadata had association. No trends with alpha diversity or beta diversity. The taxa associated with quartier (Prevotella, Oscillibacter, Collinsella, Sporobacter, Subdoligranulum follow a trend – Prevotella is inversely related to the other taxa. Sambin and Tagsyiri have the highest prevotella but lowest of the other taxa. Some papers online indicate that Collinsella is inversely related to fiber consumption, Collinsella, Sporobacter and Subdoligranulum are increased in Type2 diabetes, High Fat diet increased Oscibillibacter in mice. So these are consistent with an inverse relationship to Prevotella (where it is directly associated with fiber intake). Looking at diets of these quartiers, Samibin and Tagsyiri had 15 of 23 individuals report eating green-seeds but rest of the quartiers had a combined 13 of 45 with green-seeds consumption. No other associations stick out. With chi square these is an association between green-seeds and quartier.

**Bioinformatics Flags**

Flags for AdapterRemoval2 for burkina faso

Usearch sort by size

**Bioinformatics Tables**

Table N: Summary of metadata

Table N: Concentration value for each sample from qPCR.

Table N: Anaerobic processing data

Table N: Samples used for re-extraction and PMA treatment

Table N: Sequencing depth per sample (16S rRNA gene sequencing)

**Metadata Survey**

**Figures**

Rarefaction