

Supplemental data

Supplemental methods

Laboratory analysis

Fecal DNA extraction, 16S ribosomal RNA gene sequencing, and bioinformatics.

For DNA extraction, buffer InhibitEX was added to each sample, centrifuged at full speed to pellet samples and 600µL supernatant was pipetted to which 25µL proteinase K, and 600µL buffer AL was added. The mixture was vortexed and incubated at 70°C for 10 min. Six hundred (600)µL of 100% ethanol was added to the lysate and vortexed. Afterwards, 600µL of lysate was transferred into the QIAamp spin column and cleaned with 500µL of buffers AW1 and AW2. Finally, DNA was eluted with 200µL of buffer AE. DNA quality was evaluated by running samples on a gel and DNA quantity by qubit quantification.

We amplified the V3–V4 variable region of the 16S rRNA gene using Illumina Miseq sequencing. After sequencing, we assembled 300bp paired-end reads using the Fast Length Adjustment of SHort reads. Using the Quantitative Insights Into Microbial Ecology (QIIME), we processed the paired-end reads by filtering reads with a quality score of >25, removing mismatched barcodes, and sequences below length thresholds. Using USEARCH v7 (64-bit), we deionized reads, filtered out chimeric reads, and clustered sequences into operational taxonomic units (OTUs) at 97% identity. OTUs were aligned using the Python Nearest Alignment Space Termination tool and taxonomy was assigned using the Basic Local Alignment Search tool against the SILVA SSURef database release v123. We generated alpha diversity indices such as the Shannon diversity index in QIIME. We used genus-level OTUs for downstream analysis.

Identification of dietary patterns and bacterial compositional patterns.

For dietary patterns, the cut-level two was stable across five, six, and seven TCs. To avoid too many TCs, we selected the lowest TCs, that is, five. The five TCs at cut-level two yielded

Supplemental data

components that were very sparse. In fact, TC3, TC4, and TC5 contained only one food group, and would be irrelevant for the purpose of dimensionality reduction. Therefore, using five TCs we progressively increased the cut-level until more than one food group was loaded on all the TCs, and subsequent increase did not load more food groups. We achieved this at a cut-level 10. Therefore, we performed TT on the correlation matrix of standardized intake of the 39 food groups for five TCs at a cut-level 10.

The stability of the TCs was by using a subsampling approach, randomly sampling 80% of the data in 100 bootstrap replications. The sign patterns for TCs were calculated. For instance, a TC with loadings of $-0.9, 0.7, 0, 0.4, -0.6$ corresponds to sign patterns $-, +, 0, +, -$. The relative frequency of the sign patterns in all subsamples was calculated. The average of the relative frequencies of the TCs over three stability runs was the estimate of the stability of the TCs.