A picture containing text

Description automatically generated

**Figure S1. Baseline gut microbiota composition of the four vendors (Beijing, Guangdong, Hunan, Shanghai) at the family level.** Bars represent individual mice. Adonis test indicates significant difference in the baseline gut microbiota composition across the four vendors (P<0.001).

A screenshot of a computer

Description automatically generated with medium confidence

**Figure S2. Effects of inulin on (A) body weight, (B) daily food intake, (C) daily energy intake, and (D) 48-hr fecal sample weight of mice receiving inulin or cellulose supplementation.** Each symbol represents the mean body weight in panel A or a single data point in panels B-D. All food intakes were converted to energy intakes by multiplying food weight and its energy density (3.8 and 3.9 kcal/g for the cellulose- and inulin-based diets, respectively). The body weight data were analyzed by ordinary one-way ANOVA (Analysis of variance) with Turkey post hoc test between inulin and cellulose group. \* *P* < 0.05.

A picture containing text

Description automatically generated

**Figure S3. Dynamics of (A) Pielou’s evenness and (B) richness (number of observed ASVs) following inulin intervention.** Lines represent mean values across mice within the same vendor and shading areas represent standard error of the mean.

A screenshot of a video game

Description automatically generated with low confidence

**Figure S4.** Relative abundance of gut microbiome genes following inulin intervention. Results are shown for three vendors (Guangdong, Hunan, Shanghai) selected for metagenomic shotgun sequencing. **A**. High-dimensional gene family composition represented by robust PCoA (principal coordinate analysis) plot. R2 and P-value were obtained from Adonis analysis, which tests for the difference in gene abundances among three representative timepoints during intervention (day 0: baseline, day 5: short-term response, day 31: long-term response). **B**. Relative abundance of inulinases/fructanases, calculated as the sum of reads mapping to individual CAZy genes (GH32, GH91 and CBM38). Each dotted line represents an individual mouse. \*: P < 0.05; \*\*: P < 0.01; \*\*\*: P < 0.001.

A screen shot of a video game

Description automatically generated with low confidence

**5**D from

**A picture containing text

Description automatically generated**

**Figure S6. Reconstructed time series (lines) of bacterial load and three major short-chain fatty acids concentration by sequential Non-negative matrix factorization.** Dots represent observations, i.e., the original time series data from which reconstructed time series were built. Lines and dots are color-coded on a per-mouse basis.

Graphical user interface, application

Description automatically generated

**Figure S7.** **Bacterial taxa with significant difference in relative abundance between the inulin group and the cellulose group.** Relative abundance changes were calculated between day 0 and day 1 (A), day 0 and day 5 (B), day 0 and day 31 (C). *P*-values were obtained from Wilcoxon rank-sum test after multiple test correction via false discovery rate (FDR) estimation. \*, FDR < 0.05; \*\*, FDR < 0.01; \*\*\*, FDR < 0.001.

A screenshot of a video game

Description automatically generated with medium confidence

**Figure S8. Dynamical responses of the five inulin degraders and two generic responders.** Lines/dots: absolute abundance averaged across mice from the same vendor. Shading area: standard error of the mean values.

A screenshot of a computer

Description automatically generated with medium confidence

**Figure S9. Receiver operating characteristic (ROC) curve analysis of the similarity between training and testing datasets.** A Random Forest (RF) classifier trained to discriminate the two datasets outputs area under the ROC curve (auc) as a similarity score. ROC curves were obtained by computing the probability of samples in the full datasets (both training and test sets) predicted as being taken from the training distribution. Specificially, we first concatenated the training and test sets and assign labels 1 and 0 respectively. The new combined dataset was then stratified into 20 folds and each time, a RF classifier was trainined on 19 folds and the used to predict the probability of the remaining fold being sampled from the training set. The results clearly shows that the train and test subsets are nearly fully distinguishable from each other (auc close to 1) or very much alike (auc close to 0.5) when the full dataset was split by the “extrapolation” or the “interpolation” approach, respectively. See the main text Fig. 5B and associated texts for the details of the two train-test data split strategies.

**Background pattern

Description automatically generated with low confidence**

**Figure S10. Poor performance of Random Forest (RF) regression model in predicting short-chain fatty acids (SCFAs) concentration (see Fig. 5C of the main text for the results) cannot be rescued by using (A) alternative predictors, (B) alternative regression models, and (C,D) weighting of training samples. A**. Prediction accuracy of a RF model trained on different taxonomic- (ASV, Species, Genus, Family) or functional- (Gene, Pathway, Phenotype) predictors. The abundances of genes, pathways and phenotypes were predicted using PICRUSt2. **B**. Prediction accuracy of the MelonnPan algorithm [45] trained on the same predictors as used in panel A. Notably, MelonnPan predicts relative profiles of SCFAs from relative abundance of gut microbiota. **C**. Weights assigned to the training data. The gut microbiota composition of all samples was shown in a reduced two-dimensional UMAP (Uniform Manifold Approximation and Projection) space [49]. The bigger the weights, the larger circle sizes. Following the same approach as described in Fig. S9, we obtained the probability () of each sample in the full dataset being predicted as taken from the training subset. The weight assigned to sample was then given by , which makes intuitive sense: the higher the numerator and the lower the denominator, the closer the sample gets to the high-density regions of the test data. **D**. Prediction accuracy of an RF model built from weighted training data. The absolute abundance of bacterial taxa (grouped by the lowest classified taxonomic level) was used as predictors.

**11Mcirobiome compositional a**interventionTemporal shifts in alpha d of gut microbiotaT compositionshown robust

A picture containing dark

Description automatically generated

**Figure S12. Reduced 2-dimensional representation of the resistant starch-induced responses in bacterial load (A, corresponding to Fig. 3B) and three major SCFAs (B, corresponding to Fig. 3C).** The same figure legend applies as in the main text Fig. 3B, C.

A screenshot of a video game

Description automatically generated with medium confidence

**Figure S13. Temporal shifts in relative abundance of Bifidobacterium-related taxa—unclassified Bifidobacterium and Bifidobacterium choerinum—following inulin and resistant starch intervention.**

A picture containing text

Description automatically generated

**Figure S14. The between-vendor difference of pre-to-post changes in propionate concentration using day 5 (left panel) or day 31 (right panel) as the intervention endpoint. The P-values were obtained by** Permutational Multivariate Analaysis of Variance (PERMANOVA) with Minkowski distance as the distance metric**.**

Graphical user interface

Description automatically generated

**Figure S15**. **The microbiome-metabolome relationship is complex and time-dependent.** Dynamics of gut microbiota composition (x-axis) and total SCFA concentration (y-axis) plotted on the same graph. We used the first principal coordinate score from robust PCoA (principal coordinate analysis) ordination to represent the major changes in the gut microbiota composition (relative abundance) along the direction of maximum variance. Note that SCFAs were substantially produced between day 0 and 1 while gut microbiota composition only changes slightly. Points represent the mean PCoA coordinate score across mice from each vendor and error bars represent the standard error of the mean.**Text

Description automatically generated**

**Figure S16. Rarefaction analysis of 16S rRNA amplicon sequencing data.** Rarefaction curves were generated using the iNEXT package [81]. Solid lines represent the observed alpha diversity with the number of reads sampled, and dashed lines represent the extrapolation of the solid lines until 25% more reads. To avoid sample-to-sample bias due to variable sequencing depth (different number of reads per sample), all samples were rarefied to 38,980 sequences (black dashed line) per sample before downstream analysis.