**Baseline microbiota modulates prebiotics-mediated ecological dynamics of gut microbiota and SCFA metabolism**

Hongbin Liu1,\*, Chen Liao2,\*, Jinhui Tang1,, Junyu Chen1,, Chaobi Lei1,, Linggang Zheng1,, XXX, Joao Xavier2, Lei Dai1,#

1 CAS Key Laboratory of Quantitative Engineering Biology, Shenzhen Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

2 Program for Computational and Systems Biology, Memorial Sloan-Kettering Cancer Center, New York, NY, USA

\*: These authors contributed equally to this work

#: Correspondence: [lei.dai@siat.ac.cn](mailto:lei.dai@siat.ac.cn)

**Abstract**

Fiber intake indirectly promotes human health via gut microbiome and associated metabolism but the effects are highly personalized. While many factors are possibly at play, the well-documented individuality has been mostly associated with baseline microbiota before dietary intervention. However, little is known about the fiber-induced ecological dynamics of gut community and its baseline dependence, thus precluding in-depth understanding of the end-point response heterogeneity. Here we supplied inulin to isogenic mice that carry 4 distinct baseline microbiota for 4 weeks and longitudinally profiled microbiome and short-chain fatty acids (SCFAs) in absolute abundance. Despite the microbiota and SCFAs share consensus two-phase responses—short-term rapid stimulation and long-term partial recovery, we confirmed that the dynamical shifts in the total bacterial load, butyrate and propionate concentrations are indeed baseline-dependent. By combining statistical inference and ecological modeling, we revealed a subcommunity of 7 bacterial responders of inulin that almost fully determine the dynamics of microbiota and their relative profiles were largely controlled by baseline abundance and interspecies competition. Despite a challenge in accurate prediction of SCFAs from gut microbiota, we provided statistical evidences that the previously found correlations of baseline microbiota or total bacterial load with propionate may be mediated by these keystone bacteria. Finally, we demonstrated that some major findings from inulin are reproducible and generalizable to resistant starch intervention. Our study emphasizes the importance of ecology and quantitative modeling in mechanistic understanding of individualized responses to dietary fiber.

~~Gut microbiome is a complex ecosystem. Prebiotics such as dietary fibers can lead to system-wide shifts in microbiota composition and metabolic capacity. There is growing evidence that the response to prebiotics intervention is dependent on the baseline gut microbiota, yet the underlying dynamical process remains largely unclear. To understand the dynamical response of gut microbiota to prebiotics, we treated isogenic mice of distinct baseline gut microbiota with dietary fibers (inulin, resistant starch) and tracked the dynamics of fecal microbiome and short-chain fatty acids (SCFAs) over four weeks. Across different baseline microbiota, inulin induced a larger increase in SCFAs than resistant starch. Moreover, we observed a general pattern of bi-phasic response in microbiota diversity and SCFAs, which changed rapidly in the short term and then returned to the steady-state level. Using the generalized Lotka-Volterra model, we identified a small group of bacterial taxa as inulin responders, which collectively shaped the ecological dynamics in response to dietary shifts. We found that the dynamics of ~20% bacterial taxa was dependent on the baseline microbiota composition. Finally, we showed that the strong heterogeneity in baseline gut microbiome posed a great challenge to the prediction of SCFA profiles from gut microbiota, in accordance with previous studies in human. Our study reveals that the baseline microbiota composition modulates the dynamical response to prebiotics, calling for further studies on the role of gut microbiome in personalized nutrition.~~

**Keywords:** gut microbiome, dietary fiber, ecological dynamics, short-chain fatty acids, baseline-dependent dependence, generalized Lotka-Volterra

**Introduction**

Fermentable dietary fibers such as inulin and resistant starch are edible carbohydrate polymers that can escape colonic digestions in the small intestine but are fermented by gut microbiota in the large intestine. Human gut microbiome harbors hundreds of trillions of microbes and their associated metabolism that co-evolve with us. The major anaerobic fermentation products of gut bacteria in the large intestine are short-chain fatty acids (SCFAs). SCFAs, mainly acetate, propionate and butyrate, have broad impacts on human host physiology (e.g., stimulate glucagon-like peptide-1 secretion), immune system (e.g., regulate expression of pro-inflammatory cytokines) and intestinal homeostasis (e.g., serve as energy sources for colonocytes) [1]. Converely, impaired SCFAs production has been associated with gut microbiota dysbiosis and a wide range of dieseases [2][3][4]. To increase intestinal levels of SCFAs and eventually improve overall digestive health, some fiber molecules have been considered as “prebiotics” and made commercially available based on the premise that they can selectively enrich beneficial SCFA-producing bacteria. To test this hypothesis, diets supplemented with prebiotics have been administered as a therapeutic strategy to modulate and restore intestinal gut microbiota and SCFA levels in clinical trials.

To date, it has been well established that prebiotics can rapidly alter gut microbiota composition, including both relative abundance of individual bacteria and their total load. However, their efficacy in stimulating SCFAs remains unclear yet. Several clinical trials of healthy adults have collectively revealed that the ability of prebiotic fibers to produce SCFAs varies among individuals [9-12]. For example, Baxter et al. shows that resistant starch was able to promote butyrate production in only 63% participants [10]. Although the individualized response can arise from a combination of factors such as genetics and diet history, there is growing evidence that baseline gut microbiota is a critical factor in explaning the observed heterogeneity at multiple levels [12, 13]. The reason for the inter-individual variation in fiber consumption can be explained by the presence and absence of key fiber-degrading bacteria. One famous example is the degradation of resistant starch by *Ruminococcus brommi* and the impaired ability of healthy human donors’ fecal samples to ferment resistant starch can be *in vitro* restored by their co-incubation with *R. brommi*. The baseline dependence of microbiome and metabolome variations was much less understood, although a few studies have attributed this individuality to the lack of specific keystone species (e.g., SCFAs producers) in some baseline microbiota or the differences in their principle components. The variations in gut multi-omics can further propogate to human host, resulting in heterogenous clinical outcomes such as body mass index and glucose tolerance.

As externally provided substrates, intake of dietary fibers perturbs ecological niche of substrate utilization in the gut and thus selects for specific shifts away from the baseline composition. We propose that the compositional shift in time can be abstracted by the dynamics of a baseline-specific ecological network selected by the infleunces of dietary fibers on bacterial fitness and interactions (Fig. 1A). Several robust characteristics of ecological responses of gut microbiota have been found so far. First, consumption of fibers selectively enriches a fairly limited number of bacteria which rapidly expand and dominate the gut microbiota after substantial induction. In healthy people, *Bifidobacterium* species are one of few bacteria whose abundance consistently increases following intake of diverse fibers. Second, competitions shape the relative profiles of these key bacteria. By administering multiple fibers to the murine gut, Patnode *et al.* identified competitive inhibition as the ecological mechanism for consistent dominance of *Bacteroides cellulosilyticus* over *Bacteroids vulgatus* even though both species contain fiber-processing polysaccharide utilization loci (PULs). Third, production of SCFAs, especially butyrate, involves cross-feeding cooperations among specialized gut bacteria (Fig. 1A). By hydrolyzing complex polysaccharide fibers, primary degraders release into the gut partially breakdown products (e.g., mono- and oligo-saccharides) and fermentation metabolites (e.g., pyruvate), which can respectively benefit the secondary fiber degraders and SCFAs producers [5, 6, 7, 8]. Despite these invaluable insights, our knowledge about the fiber-induced ecological dyanmics and its dependence on baseline microbiota composition remains limited.

I

n this study, we used age- and gender-matched isogenic mice that harbor distinct baseline gut microbiota composition to study its ecological dynamics in response to dietary interventions and the inter-individual variation within these dynamical trajectories [19]. Murine models, as a complimentary alternative to clinical trials on humans, has better controls over confounding variations between individuals driven by genetic, environmental and other host differences. We monitored temporal shifts in the absolute abundance of gut bacteria and genes (quantitative PCR, 16S rRNA amplicon sequencing, metagenomics sequencing), SCFAs concentration (targeted metabolome) as well as physiological changes following the intervention of two fermentable fibers (inulin, resistant starch from maize) and cellulose (control group) (Fig. 1B). Both inulin and resistant starch are degraded by gut bacteria in the cecum and colon [20, 21], and thus have been harnessed to stimulate the production of SCFAs [10, 22]. Motivated by the hypothesis that individualized dynamics of gut microbiota and SCFAs are primed by the baseline microbial community (Fig. 1A), we developed two computational approaches that employed time series data to analyze the heterogeneity in dietary responses and study its relationships with baseline composition. As the major findings, we showed that the total bacterial load and propionate concentration exihibit baseline-dependent responses which can be linked to the abudnances of a few fiber-responsive bacteria in the baseline community. The abundances of these key bacterial responders constitute the majority of gut community and their short-term responses are controlled by their ability of fiber degradation and fiber-induced interspecific competition among them. Our study indicated that ecological inference of keystone bacteria from longitudinal data is critical for understanding the differential responses of gut microbiota to dietary fiber perturbations.

**Results**

**Heterogenous baseline gut microbiota composition across different vendors**. Age- and gender-matched isogenic mice were purchased from four commercial vendors (Beijing, Guangdong, Hunan, Shanghai), i.e. independent breeder sources. All mice were fed with cellulose-based diet 7 days prior to dietary fiber intervention. Consistent with previous studies [23, 24], these mice can be naturally divided by vendor sources into groups with distinct microbiota composition. Beta-diversity (between-sample distance) calculated by Aitchison distance shows that the baseline compositions of our mice cluster by vendor (Adonis, *P* < 0.001) and are characterized by distinct bacterial taxa (Fig. 1C, S1). For example, Shanghai mice have low relative abundances of several commensal polysaccharide-degrading bacteria such as *Muribaculaceae* and *Rikenellaceae* [25, 26]. The profound inter-vendor differences are also noticeable at the level of presence and absence of bacteria: ~65% taxa were entirely absent in at least one vendor and only ~10% bacterial taxa were present in all mice (and thus all vendors) (Fig. 1D). Due to the high between-vendor variation, mice from the same vendor can be effectively considered as independent biological replicates for each baseline microbiota composition.

A picture containing diagram

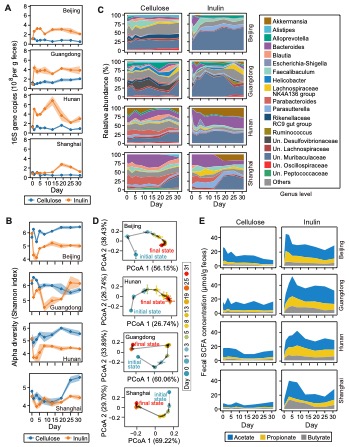
Description automatically generated

**Figure 1. High-resolution temporal profiling of murine gut microbiota and metabolome to study their individualized responses during dietary fiber intervention.** **A**. An ecological perspective of baseline-dependent dynamical responses of gut microbiota and SCFAs. Administration of dietary fibers alters ecological substrate niche in the gut and selects for an unique ecological network for each baseline microbiota type. Within the network, a few gut bacteria playing key metabolic roles as primary/secondary degraders and SCFAs fermenters drives heterogenous responses of bacteria and SCFAs via diverse ecological interactions (e.g., resource compeitition and cross-feeding). **B**. Experimental design. All mice from the four vendors were continuously fed with either dietary fiber (inulin or resistant starch)- or cellulose-supplemented diets for 4 weeks. Gray dots indicate the days on which data were collected from fecal samples. **C**. Baseline microbiota composition shown in robust PCoA (principal coordinate analysis) biplot. Isogenic age- and gender-matced mice were purchased from four different vendors (Beijing, Guangdong, Hunan, Shanghai). Gray arrows represent the dominant bacterial taxa in these samples. Adonis analysis was performed to test for differences in baseline gut microbiota composition across the four vendors (P<0.001). Un.: unclassified/uncultured. **D**. Top 4 panels: pwhite blocks indicate absence and abundance (colored blocks) the samples. Bottom panel: tall ) or ).

**Two-phase dynamics of gut microbiome and SCFA metabolism**. Dietary intake of inulin increases mice body weight gradually over time, but the amount of body weight gain is generally insignificant compared to mice treated with the cellulose control (Fig. S2A). Although there were no obvious temporal patterns in the inulin intake (Fig. S2B, C) and fecal weight (Fig. S2D), the fecal weight was dramatically lower in inulin-fed mice, suggesting that intestinal absorption of inulin may cause increased colonic transit time and decreased defecation frequency. Except for Shanghai mice, the inulin absorption nourished gut microbes by rapidly (in a day) boosting their density to more than 70% of maximum load (Fig. 2A). More interestingly, inulin induced a qualitatively consistent two-phase dynamics of microbiota (Fig. 2B)rapidly in the short-terma in the long-term (i.e., undershoot)CTinitial loss of itecological Notably, the long-term recovery is only partial for Beijing and Hunan mice with much lower gut microbiota diversity at day 31 compared to day 0.

In addition to phase-dependent responses, we also found strong tendency of gut microbiota composition to stabilize under sustained stimulation of inulin (Fig. 2D). Regardless of the baseline microbiota, the steady-state compositions are distinct from their baselines and thus represent new equilibria sustained by inulin intake. Specifically, the gut microbiota in Guangdong mice was highly resilient to inulin intervention and almost returned to its baseline composition at day 31. To quantify the rate of stabilization, we fit a harmonic oscillator model (Eq. 1-2) from physics to time series of biodiversity and calculated damping ratio—a dimensionless measure of whether and how dynamical systems approach new steady states upon perturbations—from the best-fit parameters (see Methods, Table S1). The model fits the data reasonably well with mean R2 of 73% (Table S1). The damping ratios are all equal or greater than 1, meaning that their responses were critically or overdamped and no oscillations were involved in the adaptive responses. The oscillation-free property is critical for ecosystems to maintain dynamic stability and integrity during adaptation. Although the damping ratio varies among individual mice, the dynamical responses of 61% of our mice were critical damping which leads to the fastest non-oscillatory convergence towards stationary equilibria. Interestingly, the critical damping behavior was also found in human gut microbiome recovery after transient antibiotic exposure [31].

Supporting the time-dependent shift of gut microbiota composition, metagenomic sequencing revealed that the initial (day 0), short (day 5)- and long (day 31)-term microbiomes have distinct gene family profiles (Fig. S4A). Particularly, for inulin hydrolysisS4B Consistent with the shifts in gene family composition, we observed a strong stimulating effects of inulin on the metabolism of three major SCFAs and valerate (Fig. 2E, S5). The mean peak-to-baseline concentration ratios of total SCFAs are 3.3, 3.9, 4.5 and 4.2 for Beijing, Guangdong, Hunan and Shanghai mice respectively. Since SCFAs are metabolites produced by colonic bacterial fermentation of inulin, we expect a similar phase-dependent dynamics of fecal SCFAs concentration. Indeed, both total and the three major SCFAs show two temproal phases (Fig. 2E, S5): their levels peaked in short-term before gradually decreasing until steady states (i.e., overshoot), with an exception of Shanghai mice whose propionate production was notably delayed and compromised. Despite reduced SCFAs in the second phase, the mean concentraitons of total SCFAs at day 31 still remain 60%-65% of its peak levels and 2.0x-3.5x its baseline levels. We confirmed that the long-term decline in SCFAs was not a result of reduced diet intake as the intake rate remains largely unchanged over time (Fig. S2B, C).



**Figure 2.** **Inulin-induced temporal changes in murine gut microbiota and SCFA metabolism. A.** Tload. **B.** Alpha diversity. **C.** Bacterial composition at the genus level. **D.** Bacterial composition represented by robust PCoA (principal coordinate analysis) plot. Initial and final states represent the microbiota compositions at day 0 and day 31 respectively. **E.** Fecal concentration of three major SCFAs. Beijing, Guangdong, Hunan, Shanghai are four different mice vendors. Lines (panels A,B), dots (panel D), stacked bands (panels C, E) represent the mean values across mice fromthe same vendor, and shading areas (panels A,B), error bars (panel D) represent standard error of the mean. Un.: unclassified/uncultured.

**total bacterial load and two major SCFAs** had delayed increase in bacterial load2A and produced low levels of propionate (Fig. 2E) in response to inulinThe distinct behavior of Shanghai mice indicated that the responses of total bacterial load and SCFAs may depend on the vendor and its associated baseline microbiota. test if they (or any other quantity) exihibit baseline-dependent responsesdeveloped a novel framework that separately tests for the significance of two orthogonal concepts—“baseline dependence” and “responsiveness”—based on the longitudinal data from intervention and control group (see Methods). As shown in a schematic diagram (Fig. 3A), both groups of time series are first projected onto a 2-dimensional space by sequential non-negative matrix factorization with the temporal patterns preserved after dimenstionality reduction (Fig. S6). Using the coarse-grained data representation, we then obtained two *P*-values by comparing the differential responses between the intervention and control group (“responsiveness”, *Pr*) as well as those between the four vendors (“baseline dependence”, *Pb*) in the intervention group. We determined a quantity has a baseline-dependent response if both *P*-valus are smaller than 0.05. Using this approach, we confirmed that the temporal responses of total bacterial load (Fig. 3B) and propionate (Fig. 3C) are indeed baseline-dependent. We further applied it to acetate and butyrate: only the latter metabolite shows significant baseline dependence (Fig. 3C).

A picture containing indoor, dark

Description automatically generated

**3A.**A schematic diagram of our quantitative framework to test for the significance of baseline-dependent response. The framework involves two steps: (1) projecting all time series from both intervention and control groups onto the same 2-dimensional space and (2) performing two separate statistical tests using the projected data. **B**,**C**. The 2-dimensional representation of the inulin-induced responses in the bacterial load (B) and three major SCFAs (C). Each symbol in the 2-dimensional space represents a mouse (dots: cellulose group, crosses: inulin group) and all mice data from the same vendor under the same treatment (inulin or cellulose) was used to fit an eclipse (ellipse’s radius was determined by 2 standard deviations). For each vendor, an arrow was drawn from the eclipse center of the vendor under cellulose treatment (standardized to the origin) to that under the inulin treatment.

**Inference of inulin responders and putative degraders.** The computational assay we described in the proceeding section provides a statistical way of defining “dietary fiber responders”—*P*-value for the test of “responsiveness” is smaller than 0.05 after multiple test correction. We found a total of 37 bacterial taxa with significant dynamical responses and these responders include bacterial species well known for inulin degradation, such as species of the genus *Bacteroides*. However, the ability to ferment fibers, which is testable by *in vitro* growth experiment, is only a sufficient (but not necessary) condition for the *in vivo* selectivity of fiber-induced bacterial expansion, which also requires fitness advantage over other members of their own community during resource competition. Since the inference of inulin degraders needs to take bacterial interactions into accounts, we developed a data-driven approach based on the generalized Lotka–Volterra (gLV) equation to identify putative degraders that initialize the utilization of inulin (Fig. 4A, see Methods). The gLV model assumes that fiber degradation/utilization boosts bacterial growth by an increment after controlling for the confounding growth benefits indirectly acquired from other bacteria. To estimate the uncertainty associated with model parameters, we further formulated the gLV-based inference problem in a rigorous Bayesian framework which outputs posterior distributions, rather than point estimates, of these parameters. In this framework, any bacteria with a significant positive is considered a candidate of inulin degrader.

The gLV-based ecological model identified five taxonomically heterogenous candidates for inulin degradation (Fig. 4B) , all of which have been previously inferred as the responders. Ranked from highest to lowest signal strength, these putative degraders are *Bacteroides acidifaciens* (species), *unclassified Muribaculaceae* (family), *unclassified Faecalibaculum* (genus), *unclassified Parasutterella* (genus), and *unclassified Bacteroides* (genus). Genetic or *in vitro* evidences have been found, except for *unclassified Parasutterella*, to support their functional roles in inulin degradation (Table S2). For example, members from *Bacteroides* and *Muribaculaceae* contain PULs with a *susC/susD* homologous gene pair that facilitates sensing and import of inulin [35, 36]. Putative inulin PULs were also detected in the metagenome-assembled genomes of *B. acidifaciens* and *un. Muribaculaceae* (Table S3). *Akkermancia muciniphila* and *Bacteroides uniformis*—two bacterial species whose sthe /or7—were inferred as responders but not degraders (Fig. 4B). Other than inulin-induced growth, our gLV model also predicted intense competitions among the five degraders as well as a significant positive interaction from *B. acidifaciens* to *A. muciniphila* (Fig. 4C).Considering *A. muciniphila* does not grow well *in vitro* by supplementing inulin [38], its strong but delayed response may be indirectly mediated by consuming metabolites released by the primary degraders such as *B. acidifaciens*.

**Inulin responders drive the baseline-dependent dynamics of gut microbiota.** Compared to taxonomic groups, ecological groups (eco-group) made by bacteria that perform similar ecological functions are more fundamental units for microbiome data analysis [39, 40]. In line with the notion, we divided the entire gut microbiota into three eco-groups: (1) the five putative inulin degraders; (2) the 37 inulin responders excluding the degraders (referred as generic responders below); (3) the non-responders, and showed their group-level dynamics in Fig. 4D. The group of inulin degraders dominates the total bacterial load in short-term (Guangdong, Hunan), long-term (Shanghai), or even the entire period of study (Beijing). Interestingly, both Hunan and Shanghai mice showed individualized responses of several inulin responders and the reason of individuality is linked to their baseline levels (Fig. 4D). For example, the abundances of *A. municiphila* and *B. uniformis* increased only in Hunan mice which also contain highest abundance of the two species in their baseline samples (dark yellow boxes in Fig. 4E). On the other hand, the extremely low baseline abundances of *B. acidifaciens* and *un. Muribaculaceae* in Shanghai mice (violet boxes in Fig. 4E) may explain their own sluggish responses. Spearman correlation confirms that the time-averaged absolute abundances of *B. uniformis* (P<0.001) and *B. acidifaciens* (P=0.002), but not *A. municiphia* or *un. Muribaculaceae*, are sigificiantly correlated with their baseline levels.

Although inulin stimulated growth of all five inulin degraders, their relative profiles did not remain constant and distinct temporal patterns have been observed (Fig. 4E, S8). *B. acidifaciens* and *un.* *Facelibaculum* showed transient dynamics with quick rise and drop in their absolute abundances, while the abundance of *un*. *Muribaculaceae* increased steadily and remained high throughout the entire period of study. Compared to the three degraders, *un. Parasutterella* and *un. Bacteroides* have relatively low abundances and showed complex patterns. Since the total bacterial load reached local maximum rapidly in less than 3 days (Fig. 2A, except for Shanghai mice), these degraders may compete for inulin soon after the gut microbiota has reached its temporary carrying capacity. The competition for inulin among the degraders was supported by the gLV-based inference of pairwise interactions, where *un.* *Muribaculaceae* negatively impact growth of *B. acidifaciens* and *un.* *Facaelibaculum* (Fig. 4C). This stimulation-then-competition mechanism found *in vivo* was also observed in a synthetic community of two auxotrophic yeast strains under high amino acid supplementation [42].

**Graphical user interface

Description automatically generated**

**Figure 4. Inulin** **responders/degraders shape gut microbiota dynamics in a baseline-depednent manner. A.** Generalized Lotka-Volterra (gLV) model combined with Bayesian statistics to infer inulin degraders and bacterial interactions. The gLV model summarizes the underlying ecology by three terms that additively determine bacterial growth rates: the basal growth rates (), the influences from other bacteria (), and the impacts of dietary fiber (). A primary degrader is determined when 95% credible interval of the posterior distribution of is completely to the right of 0. **B**. Posterior distribution of for five primary degraders (violet) and two generic responders (dark yellow). Generic responders are those bacteria showing statistical significance of inulin-induced response (i.e., responders) but not inferred as primary degraders. Bacterial taxa are ranked accorindg to their posterior mean of . **C**. Core ecological interaction network composed of six bacterial taxa shown in the panel B. Point and blunt arrows represent positive and negative interactions respectively. The arrow thickness is proportional to the posterior mean of the corresponding interaction coefficient. **D**. Ecological dynamics of primary degraders, generic responders (presented with two subgroups) and non-responders. **E**. Mean baseline absolute abundance of the seven bacterial taxa shown in the panel B. **F**. Temporal changes in the absolute abundances of the top three inulin degraders. For panels D and F, lines/dots and shading areas represent the mean and the standard error of the mean, respectively. Un: unclassified/uncultured.

**Inulin responders mediate the baseline-dependent propionate production.** Positive correlations were previously reported between total bacterial load and the three major SCFAs in obese males consuming resistant starch. Here we found similar positive associations, all statistically significant, and the propioinate concentration has highest Spearman coefficient with total bacterial load (Fig. 5A). By correlating the propionate concenrtratoin with three three ecological grousp as described in Fig. 4C, we further found the associaitions of xxxn total bacteria resulted from the inulin degraders and generic responders with more or less equal contribution (Fig. 5B-D). Interestingly, the top four individual bacteria taxa mostly associated the propionate concentration (*B. acidifaciens* and *un. Muribaculaceae* *A. municipila* and *B. unifomis*) are the exactly the same taxa linking baseline composition and gut microbiota response (Fig. SX). The baselinelve of Two out of fpour taxa are significant with the propionate Spearman coefficient 0.XX for Muri and XX for Bac. Uniformic. Based on these statistical analysis, we came up with a coneptutal model that reconcile fthe ifnidng isn the literature. The baseline composition determines total bacterial load based on the initial abundance of fiber degraders. Since these degraders can also produce propionate (i.e., baseline-depednent effect of propionate production), the correlations between total bacterial load and propionate elvele turn out to be indirect (Fig. 5E).

Predicting SCFAs from microbiota has been of long-standing interest but controversies regarding the predictability has not been resolved [44-46]. We evaluated the performance of machine learning models to predict the fecal SCFA concentrations using absolute abundance of bacterial taxa as predictors. All mice were split into training and test sets using two data-split approaches (Fig. 5A). The “interpolation” strategy generated balanced distribution of baseline microbiota composition between the training and test sets by randomly selecting a single mouse from each vendor as test data and leaving the other mice of the same vendor for training. By contrast, the “extrapolation” strategy produced highly unbalanced microbiota distribution by randomly selecting all mice from a vendor as test data and leaving the other vendors for training. Using Random Forest classifiers (see Methods), we confirmed that Despite a good R2 score on the training data, a Random Forest (RF) regression model generalizes poorly to the test data, especially in the “extrapolation” scenario (negative R2 means our model generalizes worse than training data average). or adding weights to training samples (Fig. 10C).

the training accuracy in terms of R2 is at least 66% regardless of SCFAs and data-split strategy. In the scenario of “interpolation”, the prediction accuracy of the three SCFAs varies from 10% to 45%, where butyrate is the most predictable SCFA probably due to its main production pathway is retained to limited bacterial species [47].

We hypothesized that the poorer performance of RF model in “extrapolation” strategy was caused by the larger microbial dissimilarity between the training and test group compared to the “interpolation”, which largely stemmed from the substantial different baseline gut microbiota composition. Some of these bacteria that are missing in at least one vendor have been inferred as SCFAs producers (Fig. S12A). Formally, we quantified the

The top porpioant predictor is un B. acidificance. The R2 for propioant in the “intraplaito” data soplit is as low as 10%, which seems to contradict the high associations between inulin degraders (either as a whole or individually() and porpinoat concentration. To find out the reason, we trained multioples RF model on different predictors using the same hyperparameter-tuning procedure. We found that the RF model is most predictivly when trained with only the five degraders, better than the model trained with all bacterial taxa. This knowledge-informed feature selection also outcompetes automatic eature selection algorithms suych as Lasso when combined with RF. Despite promising, we found that this boost is specific to propionate and the model with all bacterial taxa wins over the model with only inulin degraders for acetate and propionate. This in silico case/control eexperiment suggests that the knowledge of fiber degraders can aid feature selection and improves model prediction.

**.**

**Graphical user interface

Description automatically generated with medium confidence**

**Figure 5.** **Prediction of short-chain fatty acids concentrations from gut microbiota composition using machine learning models has limited prediction power. A**. Two data-split strategies for testing model performance. Mice in the test sets were randomly selected on a one-per-vendor basis for “interpolation” and exclusively selected from a single vendor for “extrapolation”. Data before intervention (i.e., day 0) was not included. **B**.Performances of Random Forest regression models on the training and test datasets. **D**. Receiver operating characteristic (ROC) curve analysis of the similarity between training and test datasets. A Random Forest classifier trained to discriminate the two datasets outputs area under the ROC curve (auc) as a dissimilarity metric.

**Our major findings are reproducible and generalizable.** degrader gLV-basedthe onlydegrader, probably due to the lack of quantitative dataNonetheless, t

Using the same experimental setup (Fig. 1B), we administered resistant starch to mice from the same four vendors. Consistent with previous human studies [10, 28], inulin stimulated higher production of total and three major individual SCFAs (acetate, propionate, and butyrate) than resistant starch regardless of baseline microbiota (Fig. S2B). Still, the quantitative differences in the stimulating effects of inulin and resistant starch are SCFA- and baseline-dependent. On

For resistant starch, three bacterial taxa were inferred as responders: (from highest to lowest signal strength) unclassified *Faecalibaculum* (genus), unclassised *Muribaculaceae* (family), and unclassified *Desulfovibrionaceae* (family). Some indirect evidences have been shown for the first two responders in utilizing resistant starch (Table S2). Similarly, the total relative abundance of unclassified *Muribaculaceae* and *Desulfovibrionaceae* in the baseline community explains 74% of heterogeneity in microbiota response induced by resistant starch (Fig. S14B).

In both data-split scenarios, the RF predictions are even more inaccurate using data from resistant starch intervention (Fig. S11A).

Compared to inulin, the guild of resistant starch responders plays much less dominant role in shaping the overall response of gut microbiota (Fig. S8A). In Hunan and Shanghai mice, the total density of other bacteria, despite not inferred as responders, increased sharply to higher levels than that of the responders. One possibility of this seemingly counterintuitive result might associated with the character of resistant starch, a fiber that many bacteria have the ability to utilize [12], such that no specific starch-degrading bacteria could be universally detected among four vendors. Nonetheless, we found similar trends, as observed in the inulin-group mice, for the dynamics of unclassified *Muribaculaceae* and unclassified *Faecalibaculum* (Fig. S8B): the absolute abundance of the former bacteria showed gradual but continued increase, while that of the latter bacteria showed a quick rise-and-drop pattern. GLV-based inference further suggests that the observed dynamics may be driven by mutual inhibition between the two taxa (Fig. S8C).

**Discussion**

# Discuss how our study shows that dynamics is important for understanding individualized responses

Our dynamics data confirms previous studies on the role of dietary fibers to quickly alter gut microbiota diversity and composition on the timescale of a day (largely) independent of the baseline microbiota [29, 30].

In summary, our results indicated paramount importance of dynamics for understanding inter-individual differences in dietary responses. The significance of dynamics is two-fold. Firstly, Secondly, dynamics can bridge the gap between the baseline state and the endpoint state and help identify the sources of heterogeneity. For instance, we found that the variability in bacterial load was correlated to the baseline abundance of inulin responders (*Bacteroides acidifaciens* and unclassified *Muribaculaceae)*. Taken together, characterizing dynamic responses to dietary fiber intervention across individuals, with integrated longitudinal analysis of 16S rRNA sequencing, metagenomics and metabolomics, is an important priority for microbiome research to further understanding of diet-induced responses [51]. Such studies thus have great potential to improve human health and treat gut microbiome-associated disease via microbiome engineering.

which agrees with the well-accepted notion that fiber-induced selective growth is due to competitivie interactions.

The quick response of all three degraders correspond to and may explain the initial drop of the gut microbiota diversity (Fig. 2B).

# Discuss the relevance of biphasic SCFAs dynamics to the literature

Diet-induced changes in SCFAs are often transient and vanish shortly after cessation of dietary intervention [52-55]. Our experiments add to the literature by further showing that SCFA concentrations cannot be maintained at its peak and drop by 35%-40% even under continuous inulin intake until 4 weeks. The transient responses under sustained dietary fiber intake were also observed in colorectal cancer patients [56]. Despite the drop, our data demonstrates that a continuous intervention that lasts for 31 days is sufficient to elevate and stabilize the SCFAs concentration, but it is not clear yet whether the elevated level persists after the intervention discontinues. We envision two possibilities that may explain the after-peak decrease of SCFA concentration. First, some bacteria are known to consume SCFAs and a net consumption of SCFAs may occur when SCFAs consumers dominate over producers [57]. The other possibility is increased absorption of SCFAs by host cells, leaving less produced SCFAs excreted to feces. In healthy individuals, quantification of fecal SCFAs only provides steady state levels and may not accurately reflect the level of bacterial production as most SCFAs produced in the colonic lumen (90–95%) are absorbed by the gut mucosa [58]. Nevertheless, the analysis of SCFAs in fecal samples is used as an approximation of gut levels, since excreted SCFA concentrations are associated with RS enriched diets (substrates of SCFAs-producing bacteria), inferring the relationship between intestinal SCFAs production and fecal levels [10, 59, 60].

# Discuss the advantages of our approach for baseline dependent response

Our approach has three advantages. First, the use of time series data improves our ability to distinguish responses that have different temporal trends but similar endpoint values. For example, the changes in propionate concentration from their baseline levels differ significantly among the four vendors in the short-term (day 5) but not at the endpoint (day 31) (Fig. 3B). Second, the use of control group data to assess the intervention effects, compared to the pre-to-post changes in the intervention group alone, avoids the caveat that the latter may be entirely attributed to random variations [43]. Third, the use of dimensionality reduction in our approach facilitates visualization of the variations in bacterial shifts across different baseline microbiota. For mice in each vendor, the mean response difference between the intervention and control group is represented by an arrow in a specific color. Therefore, any differences in the magnitude and direction of these vendor-specific arrows indicate baseline effects.

# Discuss how our inference approach helps tackle the replicability issue

Using public datasets, we demonstrate that our approach for inferring dietary fiber responders helps tackle replicability issue in microbiome science [61]. The structure of gLV allows for modeling bacterial interactions as covariates and accounting for confounding variables reduces the risk of producing biased estimates of indirect effects. Controlling for confounding variables is highly important for replicability between results of similar microbiome studies, as the underlying mechanisms driving microbiota dynamics is, to a great extent, shared and should play consistent roles [34], while these true microbial dynamic signatures could be largely masked by uncontrolled confounding variables [62].

Since gLV is a differential-equation model, it is naturally more amenable to analyzing longitudinal data compared to traditional cross-sectional statistical tests that based on pre-to-post changes [10, 11, 13]. Indeed, bacteria whose relative abundances were significantly altered by inulin vary depending on the day of sample collection (Fig. S5).

In fact, any scientific field that heavily relies on complex statistical analysis of large datasets may encounter challenges in replicability. Remarkably, our inference approach applied to different mouse or human gut microbiome datasets produced highly compatible results regarding inulin responders. In the mouse gut, both our data and data from Chijiiwa et al.[33] agree that the strongest inulin responder is *Bacteroides acidifaciens*, which was also reported by a third study that treated mice with inulin (data not reanalzyed) [63]. On the other hand, reanalysis of four human datasets that all administered inulin revealed that unclassified *Bifidobacterium* (in three studies) and *Anaerostipes* (in all four studies) exhibit significant responses to inulin, despite the huge differences in the pretreatment microbiota composition among human subjects in these studies (Fig. S13). Hence, consistent with a recent review that focus on the impact of inulin on human gut microbiome [64], our model further confirms the bifidogenic nature of inulin, which has been attributed to the ability of genus to efficiently take up and intracellularly degrade larger inulin. By contrast, another three inulin responders that previously associated with polysaccharide-degrading, *Faecalibacterium*, *Prevotella*, and *Lachnospiraceae*, were individually identified, suggesting that the enrichment and responses of these three bacteria might depends on the gut microbial community structure.

# Can we predict microbiota responses from baseline abundance of responders?

By combining gLV model and Bayesian regression, we identified multiple inulin and resistant starch responders from the complex microbial community of the murine intestine, which we believe play critical roles in bridging the variability in baseline microbiota to the variability in microbiota responses.

# Discuss the major reason for the failure of predicting SCFAs from microbiota composition

Considerable debates have been raised over the feasibility of predicting metabolite profiles solely based on microbiome sequencing data. For example, the MelonnPan algorithm based on linear regression accurately predicted relative abundances of >50% metabolites from metagenomes in patients with inflammatory bowel disease and healthy controls [45]. Oppositely, it was also reported that only 14% of the observed variation in the SCFAs concentration can be explained by RF regression models trained on 16S rRNA or metagenomic gene sequences in colorectal cancer patients and healthy controls [46]. While the possibility of using different regression models and patient cohorts between these controveral studies cannot be excluded, our study points out that the similarity of data distribution between training and test sets, which is strongly affected by the degree of inter-individual variability of gut microbiota, may contribute to the disagreement. Using mouse models, we showed that the predictability was completely lost when extrapolating models to predict SCFAs from gut microbiota compositions that were unseen during training (Fig. 5). In studies with humans, inconsistent model performances may be partially attributed to the neglection of stratified train-test split, leading to an undesired consequence that training and test set data are not independent and identically distributed (i.i.d.). Even though data split is stratified based on microbiome characteristics, regression models may still perform poorly in cases when inter-individual variability is too large to fulfill the i.i.d. assumption, the violation of which would cause covariate shift and machine learning to fail. To improve model predictability, large-scale human cohorts are therefore needed to cover the substantial inter-individual variation of human gut microbiome.

# Complex microbiome-metabolome relationship brings additional challenge for predicting SCFAs

Other than covariant shifts, the complex relationship between SCFAs and gut microbiota composition poses an additional challenge for predictive model development. By regressing SCFAs concentration on absolute abundance of microbes, we implicitly assume that SCFAs concentration remains unchanged if gut microbiota has a steady composition. However, this assumption does not always hold and here we provided two pieces of evidence proving that it can be violated. First, Shanghai mice showed delayed changes in total biomass (Fig. 3C) following inulin intervention, but their acetate and butyrate concentrations were neither delayed nor compromised (Fig. 2A). Second, SCFAs were highly produced by gut microbiota that maintains relatively stable composition between day 0 and day 1, regardless of vendors (Fig. S15A). It is likely the mouse gut resembles an *in vitro* culturing system at the beginning of intervention when the microbiome-metabolome relationship follows a rate model [65], i.e., gut microbiota composition determines the change in SCFAs concentration. This issue is more severe for densely sampled longitudinal data than cross-sectional data as the possibility of sampling very similar microbiota composition at two timepoints far from each other is quite low. This is consistent with previous human studies showing that short‐term diet interventions could rapidly and significantly alter gene expression of the gut microbiome, without changing the community structure [54, 66, 67]. Taken together, the decoupling of changes in SCFAs from gut microbiota responses suggests that SCFAs production may be additionally regulated at transcriptional (gene expression) and metabolically (enzyme activity) levels. Our evidence for potential molecular-level regulations justifies the emerging essentiality of integrating transcriptomics and metabolomics data in dietary response analysis.

The problem of the changed distribution between training and test dataset is formally known as covariate shift and the distribution gap can be sometimes counteracted by assigning weights to the training sets (see Methods) [48]. Unfortunately, the baseline gut microbiota between vendors seems to be too different and adding weights does not improve the prediction accuracy (Fig. S10C).

# Discuss the challenge of finding SCFAs producers

The challenges of predicting SCFAs from gut microbiota composition, as discussed above, brings similar difficulties in robust inference of SCFAs producers. Using three different approaches, we showed that the inferred top producers do not agree with each other. Taking propionate producers as an example. Using the RF model developed in Fig. 5, we ranked propionate producers based on their Gini importance scores and the top three bacterial taxa are *Bacteroides acidifaciens*, unclassified *Alloprevottella* and *Akkermansia muciniphila* (Fig. S12A). The second approach is repeated correlation [50], and we identified *Parabacteroides distasonis* as the sole significant producer (Fig. S12B), which ranked the 5th in the RF model. The third approach also trained a RF model but different variables and only subset of data were used. As discussed above, SCFAs production may follow a rate model within the first day (Fig. S15A) and we threrefore regressed the first-day changes in propionate concentration on absolute abundances of bacterial taxa. Based on Gini importance, the top three propionate producers are unclassified *Parabacteroides*, an unknown species from mouse gut metagenome, and *Parabacteroides merdae*. As shown in Fig. S15B, the absolute abundance of unclassified *Parabacteroides* showed a linear association (Pearson r: 0.71) with first-day propioinate change, while this association fades away with data after day 1 (Pearson r: 0.21). Despite the results of these three approaches have little overlap, the genus *Parabacateroides* as known propionate producers seems to regulate the proprionate level in our dataset [68].

# Limitations in translating knowledge from mice experiments to human

Studying dietary effects on gut microbiota using mouse models, as a complimentary alternative to clinical trials, has better controls over confounding variations between individuals driven by environmental and other host differences. However, we note two major limitations that may hinder translating the insights provided from our mouse experiments to humans. First, mice have much less inter-individual variability than humans [70]. Since individual mice in our study can be stratified into four distinct microbiota compositions based on vendor sources, the major conclusions we draw were essentially based comparative analysis of dietary responses among the four microbial communities each with several replicates. Whether and how well the four community types correspond to the enterotypes—distinct microbiome configurations [71] —in humans are worth further study. To soften this limitation, we have made efforts to augment the microbiome diversity by reanalyzing public dataset from a similar study. Second, the murine gut microbiota has distinct compositions from humans [72]. Although Bacteroidetes and firmicutes are two major phyla shared between mice and human, 85% genera found in the mouse gut are not detectable in the human gut [73]. For example, the family *Muribaculaceae*—a major inulin responder inferred by this study—was specific to the mouse gut [74]. Moreover, the differences in microbiota composition between mice and human diversify the gut environment at functional levels, resulting in differential dietary responses of the same microbes between the two ecosystems. For instance, the two human inulin responders—*Bifidobacterium* and *Anaerostipes*—were present in the mice gut but their relative abundances remain low and unresponsive to inulin intervention throughout the entire period of observation (Fig. S13B). This example shows that bacteria responding to dietary fiber in one of the two mammalian systems may not consistently respond in the other, suggesting that community ecology may play a selective role in the activation of fiber degraders depending on the surrounding environment.

**Methods**

**Animal experiments.** Specific-pathogen-free (SPF) female C57BL/6J mice were obtained at 6 weeks of age from four different vendors, including Beijing (A Charles River Company, Beijing, China), Hunan (Hunan Slac Jingda Laboratory Animal Company, Ltd., Changsha, China), Guangdong (Guangdong Medical Laboratory Animal Center, Foshan, China)) and Shanghai (SLAC Laboratory Animal Co., Ltd., Shanghai, China). Mice were maintained in 12-h light/dark cycle and allowed ad libitum access to food and water throughout the experiment. After acclimatizing to the diet and housing environment for 1 week, mice from each vendor were randomly separated into three groups: cellulose group (n = 5), resistant starch group (n = 5), and inulin group (n = 5). Composition of all diets including the source of dietary fibers cellulose, resistant starch, and inulin are provided in supplementary table 1 (Table S7). Fecal pellets from each mouse were freshly collected over multiple time points: day 0 (before diet change), day 1, 3, 5, 8, 13, 19, 25, and 31 (Figure 1A). Fecal samples were snap-frozen in liquid nitrogen and stored at −80 °C until further processing. At every cage change (moving the mice to a new clean cage with fresh bedding twice in one week), body weight was individually measured, and food intake and fecal output of each cage mice during the past three days per cage were measured. This study was approved by the Institutional Animal Care and Use Committee (IACUC) of the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences.

**Quantification of fecal SCFA concentration by GC-MS.** The SCFAs of mice fecal samples were analyzed by GC-MS [75]. For the sample extraction, 0.05 g of frozen feces were mixed with 300 µL of pure water containing caproic acid-6,6,6-d3 (CDN Isotopes, Quebec, Canada) as internal standard (IS, final concentration 20 µg/mL). After adding 1.0 mm diameter zirconia/silica beads (BioSpec, Bartlesville, OK), feces were homogenized for 20 s under 6500 rpm for three times, then incubated at 4 °C with shaking for 30 min, followed by centrifugation for 30 min at 13,000×g. Following extraction with anhydrous diethyl ether, the SCFA extract accurately transferred into a glass insert in a GC vial and capped tightly after added 5 µl of N, O-bis(trimethyl-silyl)-trifluoroacetamide and vortexed for 5 s. The mixture was kept in the GC vial and incubated at room temperature (22 °C) overnight (or over 8 h) before loading to GC/MS. The analysis of acetic, propionic and butyric acids was performed by Agilent 8890/7000D triple quadrupole GC/MS equipped with a capillary HP-5 ms capillary column (30 m × 0.25 mm × 0.25 µm film thickness) (Agilent Technologies). The analytes were quantified in the selected ion monitoring (SIM) mode using the target ion and confirmed by confirmative ions. The concentration was determined with reference to the peak side of IS.

**DNA extraction and quantification of bacterial load.** DNA of mice fecal samples was extracted using the QIAmp PowerFecal DNA kit (Qiagen, #12830–50) following standard manufacturer procedures. DNA samples were resuspended in Buffer C6 and quantitated using the Qubit fluorometer (ThermoFisher Scientific). To quantitatively assess bacterial load, total bacteria density were determined using qPCR as described recently [76]. The absolute abundance of a bacterial taxon was estimated by multiplication of its relative abundance and the total bacterial load.

**16S rRNA amplicon sequencing and shotgun metagenomic sequencing.** 16S rRNA gene sequencing was performed as previously described with modifications [77]. Library preparation was done using a two-step PCR method. During the first step of PCR, primers S-D-Bact-0341-b-S-17 (forward) and S-D-Bact-0785-a-A-21 (reverse) were used to target and amplify the v3-4 region [78], as well as to add second-step priming sites. Dual index codes were added to each sample at the second PCR step. The PCR products were purified with Agencourt AMPure XP magnetic beads (Beckman Coulter, Brea, CA, USA) and quality controlled with TapeStation (Agilent Technologies, Santa Clara, CA, USA). The final DNA concentrations of the purified products were measured with Qubit 2.0 fluorometer (Thermo Fisher Scientific). The purified products were pooled in equal molar concentrations, and denatured following the Illumina protocol. All sequencing was done in a single run, which was performed with a 250-cycle SP kit on the NovaSeq 6000 following the NovaSeq XP workflow (Illumina, USA). Blank controls (no sample added, processed routinely, n = 4) were included in the extraction process to control for contamination throughout processing.

Metagenomic sequencing was performed using fecal samples from the inulin diet group at day 0, 5 and 31. Extracted DNA sample was purified using silica-based columns. Metagenomics sequencing libraries were prepared with at least 2 μg of total DNA using the Nextera XT DNA sample Prep Kit (Illumina, San Diego, USA) with an equimolar pool of libraries achieved independently based on Qubit 2.0 fluorometer results combined with SYBR Green quantification (Thermo Fisher Scientific, Massachusetts, USA). The indexed libraries were sequenced with a 150-cycle S4 kit on the NovaSeq 6000 following the NovaSeq XP workflow (Illumina, USA).

**Bioinformatics analysis**. The 16S rRNA sequencing reads were analyzed by QIIME 2-2020.2 software [79]. Demultiplexed paired-end reads were trimmed to remove primers and low-quality bases with q2-cutadapt plugin. The trimmed sequences were denoised and joined with q2-dada2 plugin. Potential reagent contaminants were identified using decontam package based on either the frequency of the ASV in the blank control or the negative correlation with DNA concentration [80]. The generated feature table was filtered to remove ASVs present in only a single sample and remaining ASVs were used to construct a rooted phylogenetic tree via q2-phylogeny. Rarefaction curve analysis of the data obtained was used to estimate the completeness of microbial communities sampling and performed using the iNEXT R package [81]. Subsequently, in order to avoid sample-to-sample bias due to variable sequencing depth (different number of reads per sample), samples were rarefied to 38,980 sequences per sample. Rarefaction analysis showed that great majority of the bacteria species diversity and richness that could be sampled was captured by our sequencing depth (Fig. S18), indicated sufficient sequencing depth for majority of the analyzed samples. Estimated alpha diversity metrics, beta diversity metrics (Aitchison distance [27]) and Principle Coordinate Analysis (PCoA) using q2-diversity. Group significance between alpha and beta diversity indexes was calculated with QIIME2 plugins using the Kruskal–Wallis test and permutational multivariate analysis of variance (PERMANOVA), respectively. To assign taxonomy to the amplicon sequence variants (ASVs), the q2-feature-classifier basing on the classify-sklearn naïve Bayes taxonomy classifier was used with the SILVA (v.138) as reference database.

Indicate that PCOA plot used using DEICODE [27]

For metagenome analysis, raw sequencing reads were subjected to quality filtering and barcode trimming using KneadData (v0.5.4) by employing trimmomatic settings of 4-base wide sliding window, with average quality per base >20 and minimum length 90 bp. Reads mapping to the mouse genome were removed. Kraken2 was run against genome taxonomy database (GTDB\_r89\_54k) with default parameters [82]. Following classification by Kraken2, Bracken was used to re-estimate bacterial abundances at taxonomic levels from species to phylum using a read length parameter of 150. Next, the filtered sequences were assembled into contigs using metaSPAdes with default settings [83]. The gene abundance was analyzed and calculated as previously described with modifications [84]. Putative genes were then predicted on contigs longer than 200 base pairs using Prodigal under metagenome mode (-p meta) [85]. A non-redundant gene catalogue was constructed with CD-HIT using the parameters “-c 0.95 –aS 0.9” [86]. The abundance of each predicted gene was evaluated by mapping reads back with KMA algorithm and then normalized with the following equation: RPM = 1M × (mapped reads/gene length)/(sum of mapped reads/gene length) [87]. For all the predicted genes, CAZymes were annotated using hmmsearch against the dbCAN2 database V9 (e value <1 × 10−10; coverage >0.3) [88]. The domain with the highest coverage was selected for sequences overlapping multiple CAZyme domains. For all samples, short genomic assemblies (<2,000 bp) that could have biased the subsequent analysis were first excluded. Genomes were then binned using VAMB [89]. The binning results were refined based on the bin quality assessment (completeness >75, and contamination <15) of different binners from CheckM [90]. Taxonomic classification of each bin was determined by GTDB-tk [91], and subjected to prediction of polysaccharide utilization loci (PULz) using pipeline PULpy [92].

**Harmonic oscillation model.** The dynamic behavior of a damped harmonic oscillator is modeled by the following second-order differential equation

|  |  |
| --- | --- |
|  | Eq. (1) |

where is the displacement from its equilibrium position in the absence of external force , is the frictional force constant and is the spring constant. Given the initial position (fixed to data) and initial velocity (free parameter), the model can be solved analytically

|  |  |
| --- | --- |
|  | Eq. (2) |
|  |  |

where and . Parameters () were optimized using modified Powell algorithm with constraints . The best-fit values of and were then used to calculate the damping ratio .

**Ecological inference of dietary fiber responses.** The generalized Lotka-Volterra (gLV) model describes how the absolute abundance of bacterial species change over time

|  |  |  |
| --- | --- | --- |
|  |  | Eq. (3) |

where is the number of bacterial taxa, is the absolute abundance of taxon () , is the basal growth rate, represents the influence of taxon () on the growth of taxon , is the susceptibility coefficient that represents growth response to any given fiber, is a binary variable that indicates whether the fiber is administed at time . Bayesian regression techniques were used to parameterize the generalized Lotka-Volterra (gLV) model, as similarly used in Morjaria et al [93]. For each mice (), Eq. (1) can be transformed into a matrix form that incorporates all discrete time points of measurements (, )

|  |  |  |
| --- | --- | --- |
|  |  | Eq. (4) |

where and . The log-derivatives of on the left-hand side of Eq. (2) were estimated from a cubic spline interpolation. Using a simplified notation for Eq. (2), i.e., , we can incorpates data from all mice into a single regression model

|  |  |
| --- | --- |
|  | Eq. (5) |

The linear regression as described in Eq. (3) (for brevity ) can be further transformed into a Bayesian regression where and are normal distribution and standard deviation respectively. We used uninformative priors for all gLV parameters and Stan program [94] to produce posterior distributions for each parameter after “no U-turn” sampling of 10,000 samples from at least 3 independent Markov chain Monte Carlo traces. Since Stan is computationally expensive, we limited the inferences of dietary fiber responders to the top 20 bacterial species with the highest absolute abundances.

Since gLV models the absolute abundance of bacterial taxa, we (Fig. 3C) and then multiplied the density by their relative abundance to calculate absolute abundance. Considering ecological forces driving microbiome dynamics are largely host-independent [34], the time-series data from all mice were simultaneously fed into the gLV model after grouping bacteria according to their lowest classified taxonomic ranks.

**Significance test of baseline-dependent responses.** Sequential non-negative matrix factorization [95] was applied to transform all high-dimensional time series data from both intervention (inulin and resistant starch) and control group into two-dimensional space. We chose two factors because (1) reconstructed time series from the two latent factors preserve the quantitative trends of the untransformed time series sufficiently well (Fig. S17) and (2) two-dimensional data can be easily visualized. Suppose the reduced representation for the intervention group and control group are vectors and respectively, where () refers to the index of vendor and ( and ) refers to the index of mouse. For each vendor , both vectors were then standardized by subtracting the mean vector of the vendor in the control group, i.e., and . The significance test of the responsiveness (i.e., whether time series in the intervention group differs from that in the control group regardless of vendor) and baseline dependence (i.e., whether time series in the intervention group varies among vendors ) were performed separately using Permutational Multivariate Analaysis of Variance (PERMANOVA) with Minkowski distance as the distance metric. The Benjamini-Hochberg procedure [96] was used for controlling the false discovery rate in multiple test correction. The same PERMANOVA and muitiple test correction methods were used to test the baseline dependence based on pre-to-post changes (different of absolute abundance between day 0 and day 31).

**Random forest (RF) model.** Model development was run in a pipeline by combining normalization for data transformation, LASSO (least absolute shrinkage and selection operator) for feature selection, and RF regression for data fitting and prediction. The tolerance used in LASSO is 1e-5 and features whose coefficients below this threshold were discarded and not used to build RF regression model. Regarding model training, five hyperparameters were tuned using 5-fold cross validation and R2 as the scoring metric: constant that multiplies the L1 term in LASSO (1e-4, 1e-3, 1e-2, 1e-1, 1), the number of features to consider when looking for the best split in RF (square root, log2, 16%, 32%, 64%, 100% of all features), the maximum depth of the tree in RF (2, 4, 8, 16), the minium number of samples required to split an internal node in RF (2, 4, 8, 16), and the minimum number of samples required to be at a leaf node (1, 2, 4). We fixed the number of trees in RF model to 2,000.

ROC (Receiver Operating Characteristic) curves in Fig. 5D were obtained for RF classifiers that compute 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 dataset was 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. Suppose the probabilities of sample taken from the training and test set are and respectively. The weight assigned to sample was 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.

**Data availability**

Sequencing data, desposit to SRA

SCFA and other data: supplementary tables

**Code availability**

All scripts will be available on Github.

**References:**

[1]. Koh, A., et al., From Dietary Fiber to Host Physiology: Short-Chain Fatty Acids as Key Bacterial Metabolites. Cell, 2016. 165(6): p. 1332-1345.

[2]. Sanna, S., et al., Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. Nature Genetics, 2019. 51(4): p. 600-605.

[3]. Markey, K.A., et al., The microbe-derived short-chain fatty acids butyrate and propionate are associated with protection from chronic GVHD. Blood, 2020. 136(1): p. 130-136.

[4]. Nomura, M., et al., Association of Short-Chain Fatty Acids in the Gut Microbiome With Clinical Response to Treatment With Nivolumab or Pembrolizumab in Patients With Solid Cancer Tumors. JAMA Network Open, 2020. 3(4): p. e202895.

[5]. Solden, L.M., et al., Interspecies cross-feeding orchestrates carbon degradation in the rumen ecosystem. Nature Microbiology, 2018. 3(11): p. 1274-1284.

[6]. Rakoff-Nahoum, S., M.J. Coyne and L.E. Comstock, An Ecological Network of Polysaccharide Utilization among Human Intestinal Symbionts. Current Biology, 2014. 24(1): p. 40-49.

[7]. Belenguer, A., et al., Two routes of metabolic cross-feeding between Bifidobacterium adolescentis and butyrate-producing anaerobes from the human gut. Appl Environ Microbiol, 2006. 72(5): p. 3593-9.

[8]. Chia, L.W., et al., Cross-feeding between Bifidobacterium infantis and Anaerostipes caccae on lactose and human milk oligosaccharides. Beneficial Microbes, 2021. 12(1): p. 69-83.

[9]. Healey, G., et al., Habitual dietary fibre intake influences gut microbiota response to an inulin-type fructan prebiotic: a randomised, double-blind, placebo-controlled, cross-over, human intervention study. British Journal of Nutrition, 2018. 119(2): p. 176-189.

[10]. Baxter, N.T., et al., Dynamics of Human Gut Microbiota and Short-Chain Fatty Acids in Response to Dietary Interventions with Three Fermentable Fibers. MBio, 2019. 10(1).

[11]. Deehan, E.C., et al., Precision Microbiome Modulation with Discrete Dietary Fiber Structures Directs Short-Chain Fatty Acid Production. Cell Host & Microbe, 2020.

[12]. Venkataraman, A., et al., Variable responses of human microbiomes to dietary supplementation with resistant starch. Microbiome, 2016. 4(1): p. 33.

[13]. Nguyen, N.K., et al., Gut microbiota modulation with long-chain corn bran arabinoxylan in adults with overweight and obesity is linked to an individualized temporal increase in fecal propionate. Microbiome, 2020. 8(1).

[14]. Ze, X., et al., Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon. ISME J, 2012. 6(8): p. 1535-43.

[15]. Maini Rekdal, V., et al., Discovery and inhibition of an interspecies gut bacterial pathway for Levodopa metabolism. Science, 2019. 364(6445): p. eaau6323.

[16]. Maldonado-Gómez, M.X., et al., Stable Engraftment of Bifidobacterium longum AH1206 in the Human Gut Depends on Individualized Features of the Resident Microbiome. Cell Host & Microbe, 2016. 20(4): p. 515-526.

[17]. Rodriguez, J., et al., Discovery of the gut microbial signature driving the efficacy of prebiotic intervention in obese patients. Gut, 2020: p. gutjnl-2019-319726

.

[18]. Stein, R.R., et al., Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. PLoS Comput Biol, 2013. 9(12): p. e1003388.

[19]. Liu, Z., et al., Microbiome-based stratification to guide dietary interventions to improve human health. Nutrition Research, 2020. 82: p. 1-10.

[20]. Ahmed, W. and S. Rashid, Functional and therapeutic potential of inulin: A comprehensive review. Crit Rev Food Sci Nutr, 2019. 59(1): p. 1-13.

[21]. Cerqueira, F.M., et al., Starch Digestion by Gut Bacteria: Crowdsourcing for Carbs. Trends in Microbiology, 2019.

[22]. Lordan, C., et al., Potential for enriching next-generation health-promoting gut bacteria through prebiotics and other dietary components. Gut microbes, 2019: p. 1-20.

[23]. Parker, K.D., et al., Microbiome Composition in Both Wild-Type and Disease Model Mice Is Heavily Influenced by Mouse Facility. Frontiers in Microbiology, 2018. 9.

[24]. Ericsson, A.C., et al., Effects of Vendor and Genetic Background on the Composition of the Fecal Microbiota of Inbred Mice. PLOS ONE, 2015. 10(2): p. e0116704.

[25]. Lagkouvardos, I., et al., Sequence and cultivation study of Muribaculaceae reveals novel species, host preference, and functional potential of this yet undescribed family. Microbiome, 2019. 7(1).

[26]. Pereira, F.C., et al., Rational design of a microbial consortium of mucosal sugar utilizers reduces Clostridiodes difficile colonization. Nat Commun, 2020. 11(1): p. 5104.

[27]. Martino, C., et al., A Novel Sparse Compositional Technique Reveals Microbial Perturbations. mSystems, 2019. 4(1).

[28]. Rahat-Rozenbloom, S., et al., The acute effects of inulin and resistant starch on postprandial serum short-chain fatty acids and second-meal glycemic response in lean and overweight humans. European journal of clinical nutrition, 2017. 71(2): p. 227-233.

[29]. Wu, G.D., et al., Linking long-term dietary patterns with gut microbial enterotypes. Science, 2011. 334(6052): p. 105-8.

[30]. Creswell, R., et al., High-resolution temporal profiling of the human gut microbiome reveals consistent and cascading alterations in response to dietary glycans. Genome Medicine, 2020. 12(1).

[31]. Shaw, L.P., et al., Modelling microbiome recovery after antibiotics using a stability landscape framework. The ISME Journal, 2019. 13(7): p. 1845-1856.

[32]. Scott, K.P., et al., Prebiotic stimulation of human colonic butyrate-producing bacteria and bifidobacteria, in vitro. FEMS Microbiol Ecol, 2014. 87(1): p. 30-40.

[33]. Chijiiwa, R., et al., Single-cell genomics of uncultured bacteria reveals dietary fiber responders in the mouse gut microbiota. Microbiome, 2020. 8(1): p. 5-14.

[34]. Bashan, A., et al., Universality of human microbial dynamics. Nature, 2016. 534(7606): p. 259-262.

[35]. Rakoff-Nahoum, S., K.R. Foster and L.E. Comstock, The evolution of cooperation within the gut microbiota. Nature, 2016. 533(7602): p. 255-259.

[36]. Koropatkin, N.M., E.A. Cameron and E.C. Martens, How glycan metabolism shapes the human gut microbiota. Nat Rev Microbiol, 2012. 10(5): p. 323-35.

[37]. Belzer, C., et al., Microbial Metabolic Networks at the Mucus Layer Lead to Diet-Independent Butyrate and Vitamin B12 Production by Intestinal Symbionts. mBio, 2017. 8(5).

[38]. Zhou, K., Strategies to promote abundance of Akkermansia muciniphila, an emerging probiotics in the gut, evidence from dietary intervention studies. J Funct Foods, 2017. 33: p. 194-201.

[39]. Goldford, J.E., et al., Emergent simplicity in microbial community assembly. Science, 2018. 361(6401): p. 469-474.

[40]. Wu, G., et al., Guild-based analysis for understanding gut microbiome in human health and diseases. Genome Med, 2021. 13(1): p. 22.

[41]. Falony, G., et al., Cross-Feeding between Bifidobacterium longum BB536 and Acetate-Converting, Butyrate-Producing Colon Bacteria during Growth on Oligofructose. Applied and Environmental Microbiology, 2006. 72(12): p. 7835-7841.

[42]. Hoek, T.A., et al., Resource Availability Modulates the Cooperative and Competitive Nature of a Microbial Cross-Feeding Mutualism. PLOS Biology, 2016. 14(8): p. e1002540.

[43]. Atkinson, G. and A.M. Batterham, True and false interindividual differences in the physiological response to an intervention. Experimental Physiology, 2015. 100(6): p. 577-588.

[44]. Yin, X., et al., A Comparative Evaluation of Tools to Predict Metabolite Profiles From Microbiome Sequencing Data. Frontiers in Microbiology, 2020. 11.

[45]. Mallick, H., et al., Predictive metabolomic profiling of microbial communities using amplicon or metagenomic sequences. Nature Communications, 2019. 10(1).

[46]. Sze, M.A., et al., Fecal Short-Chain Fatty Acids Are Not Predictive of Colonic Tumor Status and Cannot Be Predicted Based on Bacterial Community Structure. mBio, 2019. 10(4): p. e01454-19.

[47]. Vital, M., A. Karch and D.H. Pieper, Colonic Butyrate-Producing Communities in Humans: an Overview Using Omics Data. mSystems, 2017. 2(6).

[48]. Storkey, A., When training and test sets are different: characterizing learning transfer. Dataset shift in machine learning, 2009. 30: p. 3-28.

[49]. McInnes, L., J. Healy and J. Melville, Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426, 2018.

[50]. Bakdash, J.Z. and L.R. Marusich, Repeated Measures Correlation. Frontiers in Psychology, 2017. 8.

[51]. Johnson, A.J., et al., A Guide to Diet-Microbiome Study Design. Frontiers in Nutrition, 2020. 7.

[52]. Baxter, N.T., et al., The Glucoamylase Inhibitor Acarbose Has a Diet-Dependent and Reversible Effect on the Murine Gut Microbiome. mSphere, 2019. 4(1).

[53]. Walker, A.W., et al., Dominant and diet-responsive groups of bacteria within the human colonic microbiota. ISME J, 2011. 5(2): p. 220-30.

[54]. David, L.A., et al., Diet rapidly and reproducibly alters the human gut microbiome. Nature, 2014. 505(7484): p. 559-63.

[55]. Hiel, S., et al., Effects of a diet based on inulin-rich vegetables on gut health and nutritional behavior in healthy humans. Am J Clin Nutr, 2019. 109(6): p. 1683-1695.

[56]. Nordgaard, I., et al., Colonic production of butyrate in patients with previous colonic cancer during long-term treatment with dietary fibre (Plantago ovata seeds). Scand J Gastroenterol, 1996. 31(10): p. 1011-20.

[57]. Liao, C., et al., Integrated, systems metabolic picture of acetone-butanol-ethanol fermentation by Clostridium acetobutylicum. Proc Natl Acad Sci U S A, 2015. 112(27): p. 8505-10.

[58]. Sakata, T., Pitfalls in short-chain fatty acid research: A methodological review. Animal Science Journal, 2019. 90(1): p. 3-13.

[59]. McOrist, A.L., et al., Fecal Butyrate Levels Vary Widely among Individuals but Are Usually Increased by a Diet High in Resistant Starch. Journal of Nutrition, 2011. 141(5): p. 883-889.

[60]. Haenen, D., et al., A Diet High in Resistant Starch Modulates Microbiota Composition, SCFA Concentrations, and Gene Expression in Pig Intestine. Journal of Nutrition, 2013. 143(3): p. 274-283.

[61]. Schloss, P.D., Identifying and Overcoming Threats to Reproducibility, Replicability, Robustness, and Generalizability in Microbiome Research. mBio, 2018. 9(3).

[62]. Ghosh, T.S., et al., Adjusting for age improves identification of gut microbiome alterations in multiple diseases. eLife, 2020. 9.

[63]. Then, C.K., et al., Association of Bacteroides acidifaciens relative abundance with high-fibre diet-associated radiosensitisation. BMC Biol, 2020. 18(1): p. 102.

[64]. Le Bastard, Q., et al., The effects of inulin on gut microbial composition: a systematic review of evidence from human studies. European Journal of Clinical Microbiology & Infectious Diseases, 2019.

[65]. Gurry, T., et al., Functional heterogeneity in the fermentation capabilities of the healthy human gut microbiota. bioRxiv, 2020: p. 2020.01.17.910638.

[66]. Tap, J., et al., Gut microbiota richness promotes its stability upon increased dietary fibre intake in healthy adults. Environmental Microbiology, 2015. 17(12): p. 4954-4964.

[67]. Sasaki, D., et al., Low amounts of dietary fibre increase in vitro production of short-chain fatty acids without changing human colonic microbiota structure. Sci Rep, 2018. 8(1): p. 435.

[68]. Gotoh, A., et al., Use of Gifu Anaerobic Medium for culturing 32 dominant species of human gut microbes and its evaluation based on short-chain fatty acids fermentation profiles. Biosci Biotechnol Biochem, 2017. 81(10): p. 2009-2017.

[69]. Ubeda, C., et al., Vancomycin-resistant Enterococcus domination of intestinal microbiota is enabled by antibiotic treatment in mice and precedes bloodstream invasion in humans. J Clin Invest, 2010. 120(12): p. 4332-41.

[70]. Nguyen, T.L., et al., How informative is the mouse for human gut microbiota research? Dis Model Mech, 2015. 8(1): p. 1-16.

[71]. Costea, P.I., et al., Enterotypes in the landscape of gut microbial community composition. Nature Microbiology, 2018. 3(1): p. 8-16.

[72]. Hugenholtz, F. and W.M. de Vos, Mouse models for human intestinal microbiota research: a critical evaluation. Cellular and Molecular Life Sciences, 2017.

[73]. Ley, R.E., et al., Obesity alters gut microbial ecology. Proc Natl Acad Sci U S A, 2005. 102(31): p. 11070-5.

[74]. Lagkouvardos, I., et al., The Mouse Intestinal Bacterial Collection (miBC) provides host-specific insight into cultured diversity and functional potential of the gut microbiota. Nature Microbiology, 2016. 1(10).

[75]. Zhang, S., H. Wang and M. Zhu, A sensitive GC/MS detection method for analyzing microbial metabolites short chain fatty acids in fecal and serum samples. Talanta, 2019. 196: p. 249-254.

[76]. Jian, C., et al., Quantitative PCR provides a simple and accessible method for quantitative microbiota profiling. PLOS ONE, 2020. 15(1): p. e0227285.

[77]. Gohl, D.M., et al., Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. Nature Biotechnology, 2016. 34(9): p. 942-949.

[78]. Klindworth, A., et al., Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Research, 2013. 41(1): p. e1-e1.

[79]. Bolyen, E., et al., Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol, 2019. 37(8): p. 852-857.

[80]. Davis, N.M., et al., Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome, 2018. 6(1).

[81]. Hsieh, T.C., K.H. Ma and A. Chao, iNEXT: an R package for rarefaction and extrapolation of species diversity (H ill numbers). Methods in Ecology and Evolution, 2016. 7(12): p. 1451-1456.

[82]. Wood, D.E., J. Lu and B. Langmead, Improved metagenomic analysis with Kraken 2. Genome Biology, 2019. 20(1).

[83]. Nurk, S., et al., metaSPAdes: a new versatile metagenomic assembler. Genome Res, 2017. 27(5): p. 824-834.

[84]. Zhao, Z., F. Baltar and G.J. Herndl, Linking extracellular enzymes to phylogeny indicates a predominantly particle-associated lifestyle of deep-sea prokaryotes. Science advances, 2020. 6(16): p. eaaz4354.

[85]. Hyatt, D., et al., Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics, 2010. 11: p. 119.

[86]. Fu, L., et al., CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics, 2012. 28(23): p. 3150-2.

[87]. Clausen, P.T.L.C., F.M. Aarestrup and O. Lund, Rapid and precise alignment of raw reads against redundant databases with KMA. BMC Bioinformatics, 2018. 19(1).

[88]. Zhang, H., et al., dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Research, 2018. 46(W1): p. W95-W101.

[89]. Nissen, J.N., et al., Improved metagenome binning and assembly using deep variational autoencoders. Nature Biotechnology, 2021.

[90]. Parks, D.H., et al., CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res, 2015. 25(7): p. 1043-55.

[91]. Chaumeil, P.A., et al., GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics, 2019.

[92]. Stewart, R.D., et al., Open prediction of polysaccharide utilisation loci (PUL) in 5414 public Bacteroidetes genomes using PULpy. bioRxiv, 2018: p. 421024.

[93]. Morjaria, S., et al., Antibiotic-Induced Shifts in Fecal Microbiota Density and Composition during Hematopoietic Stem Cell Transplantation. Infect Immun, 2019. 87(9).

[94]. Carpenter, B., et al., Stan: a probabilistic programming language. Grantee Submission, 2017. 76(1): p. 1-32.

[95]. Mackevicius, E.L., et al., Unsupervised discovery of temporal sequences in high-dimensional datasets, with applications to neuroscience. Elife, 2019. 8.

[96]. McDonald, J.H., Handbook of biological statistics. Vol. 2. 2009: sparky house publishing Baltimore, MD.