**Ecological dynamics of the gut microbiome in response to dietary fiber**

Hongbin Liu1,\*, Chen Liao2,\*, Jinhui Tang1,, Junyu Chen1,, Chaobi Lei1,, Linggang Zheng1,, Lu Wu1, … 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**

Dietary fibers are generally thought to benefit intestinal health. Their impacts on the composition and metabolic function of the gut microbiome, however, vary greatly from individual to individual. Previous research showed that each individual’s response to fibers depends on their baseline gut microbiome, but the ecology driving microbiota remodeling during fiber intake remained unclear. Here, we studied the long-term dynamics of gut microbiome and short-chain fatty acids (SCFAs) in isogenic mice with distinct microbiota baselines fed with the fermentable fiber inulin compared to the non-fermentable fiber cellulose. Inulin produced generally rapid, yet baseline-specific, responses followed by gradual stabilization to new equilibria. We parameterized an ecology model from the time series data, revealing a group of bacteria whose growth significantly increases in response to inulin, and whose baseline abundance and intertaxa competition explain the observed dynamics of microbiome density and community composition. The production of SCFAs, such as propionate, is associated with these inuline responders, but fecal levels of these metabolites could not be accurately predicted due to their large variation among individuals. Finally, our methods and major findings are generalizable to dietary resistant starch. This study emphasizes the importance of ecological modeling to understand microbiome responses to dietary changes and the need for personalized interventions.

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

**Introduction**

Fermentable dietary fibers, such as inulin and resistant starch, are edible carbohydrate polymers that escape colonic digestions in the small intestine and are fermented by gut microbiota in the large intestine. Anaerobic fermentation by bacteria in the large intestine produces short-chain fatty acids (SCFAs)—mainly acetate, propionate and butyrate—which have broad impacts on intestinal health and immunity [1][2][3]. Impaired SCFAs production has been associated with diesease [4, 5]. Dietary fibers can selectively enrich beneficial gut bacteria [6] and could be administered as “pre-biotic” therapies to restore intestinal gut microbiota and elevate SCFA levels [7, 8].

Previous work showed that dietary fibers cause rapid changes in microbiota compsotion and density [9, 10]. But the ability of fibers to increase SCFA production varies among individuals [11-14]. For example, Baxter *et al.* showed that resistant starch was able to promote butyrate production in only 63% participants [12]. The individualized response can arise from a combination of factors such as genetics and diet history. But the baseline gut microbiota is also a critical factor [14, 15]. The feces of some healthy human donors’ fail to ferment resistant starch, but this function can be restored by co-incubation with *R. brommi,* a well-known degrader of resistant starch [16]. Person-to-person variation in the bacterial and metabolomic composition of gut microbiome can further propogate to their human host, impacting biological variables such as body mass index [17] and glucose tolerance [18].

Dietary fibers select from the pool of baseline community for microbial taxa that can use fibers as substrates for growth, and these responders could further impact the entire gut microbial community through a baseline-dependent ecological network (#cite PMID: 26542567#) (**Fig. 1A**). The primary users of dietary fibers are relatively few in low-fiber diet, but they can rapidly expand and dominate the gut microbiota after substantial induction [19]. Production of some SCFAs, especially butyrate, involves cross-feeding cooperations among specialized gut bacteria. 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 [20, 21]. Despite increasing interests in diet-induced modulation of microbiota composition and ecology [22], a system-level, quantitative understanding of the ecological dyanmics of gut microbiome under dietary interventions and their baseline dependences is still lacking.

In this study, we profiled longitudinally the gut microbiota of mice to study the basis for the baseline-dependent dynamical response to dietary fibers. We used the time series data to infer the ecological network that explains why the microbiota fiber responses vary with their baseline composition. The derived network shows that the growth and ecological interactions of key responders—bacteria with significant *in vivo* growth in inulin or resistant starch compared to cellulose—drive the major shifts in microbiota composition. We further showed that their baseline abundances predict individualized changes in bacterial absolute abundance and propionate concentration in feces. This study provides a framework to identify the ecological drivers of microbiota response to dietary interventions, which is critical for understanding the individualized responses of gut microbiota and the design of targeted modulations.

**Results**

**Isogenic mice from different origins vary in their baseline gut microbiota**

We used age- and gender-matched isogenic mice that harbor distinct baseline gut microbiota composition to study the dynamical response to dietary interventions and the inter-individual variation in ecological dynamics [23]. Mice were purchased from four commercial vendors (labeled as Beijing, Guangdong, Hunan, Shanghai, see **Methods**), i.e. independent breeder sources. All mice were fed with cellulose-based diet 7 days prior to dietary fiber intervention. We monitored temporal shifts in the absolute abundance and community composition of gut bacteria (by quantitative PCR, 16S rRNA amplicon sequencing, metagenomics sequencing), SCFAs concentration (by targeted metabolome) as well as physiological changes following the intervention of fermentable dietary fibers and cellulose (control group) (**Fig. 1B**). The two dietary fibers used in this study, inulin and resistant starch from maize, are able to be degraded by gut bacteria in the cecum/colon [24, 25] and to stimulate the production of SCFAs [8, 12].

Consistent with previous studies [26, 27], these mice grown in vendor-specific housing and feeding conditions 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, Fig. S1A**). For example, Shanghai mice have low relative abundances of several commensal polysaccharide-degrading bacteria such as *Muribaculaceae* and *Rikenellaceae* [28, 29]. 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**). Interestingly, the prevalence and abundance of bacterial taxa in these baseline samples exihibit a strong positive linear relationship in log-log space (**Fig. S1B**). 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. Throughout the observation of our experiment, the body weight of mice gradually increased over time, but the gain in body weight is generally insignificant between the inulin treatment group and the cellulose control group (**Fig. S2A**). Among the different experimental groups, there were no obvious difference in food intake and fecal weight (**Fig. S2**).

**Baseline-dependent microbiota dynamics in response to inulin**

Inulin intervention rapidly promoted the absolute abundance of gut bacteria on the time scale of days, except for Shanghai mice (**Fig. 2A**). More interestingly, inulin induced a two-phase dynamical response in the gut microbiota diversity (**Fig. 2B**), which dropped rapidly in the short-term and recovered gradually in the long-term. The initial loss of diversity is primarily due to the changes in evenness (**Fig. S3A**), not richness (**Fig. S3B**), suggesting an expansion of certain bacterial taxa. Indeed, we observed rapid but non-monotonic changes in the relative abundances of several dominant bacterial genera, such as *Bacteroides* and *unclassified Muribaculaceae* (**Fig. 2C**). Notably, the long-term recovery of microbiota diversity is only partial for Beijing and Hunan mice (i.e. lower gut microbiota diversity at day 31 compared to day 0). By metagenomic sequencing, we observed temporal changes in the functional capacity of gut microbiome. Specificially, the initial (day 0), short-term (day 5) and long-term (day 31) microbiomes have distinct gene family profiles (**Fig. S4A**) and the relative abundance of genes encoding enzymes for inulin utilization (inulinases/fructanases) significantly increased after intervention (**Fig. S4B**). Collectively, our longitudinal profilings are consistent with previous observations that dietary fibers profoundly impact the ecology and function of gut microbiota [30,31]. In addition, the gut microbiota tended 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.

The changes in the gut microbiota were accompanied by changes in the levels of three major SCFAs (acetate, propionate and butyrate) and valerate (**Fig. 2E, Fig. S5**). 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 (acetate, propionate, butyrate, iso-butyrate, valerate, iso-valerate) and the three major SCFAs show two temproal phases: their levels peaked in short-term before gradually decreasing until steady states, with an exception of Shanghai mice whose propionate production was notably delayed and compromised. The mean peak-to-baseline concentration ratios of total SCFAs were 3.3, 3.9, 4.5 and 4.2 for Beijing, Guangdong, Hunan and Shanghai mice respectively. The long-term decline in SCFAs was not a result of reduced diet intake, as the intake rate remained unchanged over time (**Fig. S2**). Despite reduced SCFAs in the second phase, the mean concentrations of total SCFAs at day 31 were still 2.0-3.5 fold of its baseline levels.

We have shown above that Shanghai mice had a delayed increase in bacterial absolute abundance (**Fig. 2A**) and produced low levels of propionate (**Fig. 2E**) in response to inulin. The distinct behavior of Shanghai mice indicated that the responses of bacterial absolute abundance and SCFAs may depend on the baseline microbiota. To formally test this, we separately tested for the statistical significance of two orthogonal properties—“baseline dependence” and “responsiveness”—based on the time series data of intervention group and control group (see **Methods**). Time series data of both groups were projected onto a 2-dimensional space by sequential non-negative matrix factorization [33] to capture representative temporal trends (**Fig. 3A, 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 in the intervention group (“baseline dependence”, *Pb*). This approach confirmed that the dynamical responses of bacterial load (**Fig. 3B**), propionate and butyrate (**Fig. 3C**) were nontrivial and baseline-dependent (both *Pr* and *Pb* < 0.05 after multitest correction).

**Primary degraders of inulin respond fast and drive community dynamics**

We used the generalized Lotka–Volterra (gLV) model to infer key ecological drivers of gut microbiota response to dietary fiber intervention (**Fig. 4A,** see **Methods**). The gLV model assumes that degradation and subsequent utilization of dietary fibers boost bacterial growth (the amount of increment is parameterized by ). To estimate the uncertainty associated with model parameters, we formulated the gLV-based inference in a Bayesian framework which outputs posterior distributions of estimated parameters, rather than point estimates. Since our gLV-based framework incorporates bacterial interactions as confounding variables, any bacteria taxa with a significantly positive “fiber response” is considered a putative “primary degrader” of inulin. We identified five such bacterial taxa grouped at heterogeneous taxonomic levels, including *Bacteroides acidifaciens* (species), *Muribaculaceae* (family), *Faecalibaculum* (genus), *Parasutterella* (genus), and *Bacteroides* (genus) (**Fig. 4B**). Note that these taxa grouped 16S sequences classified at the specified taxonomic level excluding those classified at lower levels (see **Methods**). We further showed that the inference is robust to the method for grouping sequences (**Fig. S7**): the inferred primary degraders using our approach (grouped at the lowest classified taxonomic level) and OTU (Operational Taxonomic Unit)-picking (97% sequence similarity) are consistent with each other.

For four of the five putative inulin degraders (except *Parasutterella*), we found genetic evidence and/or *in vitro* experiments from literature to support their functional roles in inulin degradation (**Table S1**). For example, members from *Bacteroides* and *Muribaculaceae* contain PULs with a *susC/susD* homologous gene pair that facilitates sensing and import of inulin [34, 35]. Putative inulin PULs were also detected in the metagenome-assembled genomes of *B. acidifaciens* and *Muribaculaceae* (**Table S2**). Furthremore, we analyzed the data from an independent study [36], which profiled the murine gut microbiota composition after inulin intervention for two weeks. Analysis of this independent experiment found similar dynamics in gut microbiota diversity and composition (**Fig. S8A,B**). Although bacterial absolute abudance was unavailable in this data, we applied gLV-based inference to the relative abundance profiles and again identified *B. acidifaciens* as a primary degrader of inulin (**Fig. S8C**).

Since not all bacteria grow by metabolizing inulin directly, we alternatively used the statistical test of responsiveness (**Fig. 3A;** *Pr*<0.05 after multitest correction) to identify all bacterial taxa that exihibited differential responses between inulin and cellulose groups, regardless of the ecological mechanism. We found a total of 37 taxa with significant dynamical responses (**Table S3**), including the five putative primary degraders as inferred by the gLV model. Among the remaining 32 “generic responders”, *Akkermancia muciniphila* and *Bacteroides uniformis* are the most abundant taxa whose relative abundances significantly increased after day 5 (**Fig. S9, Fig. S10**). Unlike the gLV-based approach, the statistical analysis on time series data did not control for indirect effects on bacterial growth via ecological interactions, so the two “generic responders” may indirectly benefit from the primary degraders. Indeed, inference of the gLV model did not identify *A. muciniphila* as a “primary degrader”, but suggested that its growth was facilitated by *B. acidifaciens* (**Fig. 4C**).

Eco-groups (or guilds) of bacterial taxa that perform similar functions are a useful concept to understand microbial ecology [37, 38]. We divided the entire gut microbiota into three eco-groups: (1) 5 primary degraders of inulin; (2) 32 generic responders to inulin intervention; (3) non-responders. The group-level dynamics showed that the primary degraders clearly dominated the microbiota response (**Fig. 4D**). The short-term rise in the absolute abundance of a few bacterial taxa corresponds to the initial decline of the gut microbiota evenness soon after the intervention (**Fig. 2B**). More interestingly, the baseline-dependent responses can be causally linked to the initial composition of key bacterial taxa (**Fig. 4D,E**). For example, the abundances of *A. municiphila* and *B. uniformis* increased dramatically in Hunan mice (**Fig. S10**), which contained the highest abundance of these two species in the baseline (dark yellow box frames in **Fig. 4E**). Similarly, the extremely low baseline abundances of *B. acidifaciens* and *Muribaculaceae* in Shanghai mice (violet box frames in **Fig. 4E**) may explain the sluggish responses in bacterial absolute abundance and SCFAs (**Fig. S5, S10**).

Our gLV-based inference suggested strong competition among primary degraders, where *Muribaculaceae* inhibited the growth of *B. acidifaciens* and *Facaelibaculum* (**Fig. 4C**). Indeed, *B. acidifaciens* and *Facelibaculum* showed transient dynamics with a quick rise and drop in their absolute abundances, while the abundance of *Muribaculaceae* increased steadily and remained high during the entire period of observations (**Fig. 4F, S10**). Our results are consistent with previous studies by Patnode *et al.* [39] that identified competitive inhibition as the ecological mechanism for consistent dominance of *Bacteroides cellulosilyticus* over *Bacteroids vulgatus,* even though both species contain fiber-processing PULs. Taken together, we demonstrate that primary degraders and their intertaxa competitions are key drivers of the baseline-dependent ecological dynamics of microbiota response to dietary fibers.

**Baseline-dependent SCFA production and its association with gut microbiota composition**

The dynamics of SCFAs during inulin intervention varied substantially across different baselines (**Fig. S5**). Shanghai mice produced the lowest level of propionate (**Fig. S5**); these mice also showed the lowest response in bacterial load (**Fig. 2A, Fig.4D**), due to very low abundance of some primary degraders and generic responders of inulin in the baseline microbiota (**Fig. 4E**). We hypothesized that these key taxa may directly contribute to propionate production and found that the baseline abundances of *B. acidifaciens*, *Muribaculaceae*, *A. municiphila*, *B. uniformis* were positively correlated the propionate concentration (**Fig. 5A,** left panel). Indeed, *Muribaculaceae*, *A. municiphila* and *B. uniformis* have been previously found to produce propionate *in vitro* and/or *in vivo* (**Table S1**). As a result and consistent with a previous study [40], there was a strong positive association between bacterial load and propionate concentration (**Fig. 5A**, right panel; P<0.001), as the two are both baseline-dependent. In contrast, the association between bacterial load and other SCFAs was not significant (**Fig. S11**).

Given that the gut microbiota is strongly associated with the fecal levels of fiber fermentation products, we asked whether we could quantitatively predict SCFAs concentrations from the microbiota composition measured at the same time. We evaluated the performance of machine learning models to predict the fecal SCFA concentrations using absolute abundance of bacterial taxa as predictors. All mice in our experiments were split into training data and test data using different data-split approach (**Fig. 5B**). The “interpolation” approach generated balanced distribution of baseline microbiota composition between the training and test data (**Fig. S12A**), by randomly selecting a single mouse from each vendor as test data and using the other mice for training. In contrast, the “extrapolation” approach produced highly unbalanced microbiota distribution between the training and test data (**Fig. S12B**), by randomly selecting all mice from a vendor as test data and using mice of the other vendors for training. Although the Random Forest (RF) regression model fitted the training data reasonably well (R2 ≥0.66 regardless of SCFAs and data-split approaches), the predictions generalized poorly to the test data: R2 of SCFAs ranged from 0.1 to 0.45 for “interpolation” but dropped below 0 for “extrapolation” (**Fig. 5C**). We further showed that the low predictability in extrapolation cannot be substantially improved by using alternative predictors (**Fig. S13A**), models (**Fig. S13B**) or adding weights to training samples (**Fig. S13C**). Overall, we found that machine learning models based on gut microbiota composition had low or no predictive power for fecal SCFA concentration, if the gut microbiota of interest was significantly different from the baselines covered in training data. This agrees with previous studies in humans finding that the substantial inter-individual variation of gut microbiome could impede the predictive power of machine learning models [41].

**Response to resistant starch validates model developed from response to innulin**

To study whether our ecological framework can be generalized to study the dynamical responses of gut microbiota to other dietary fibers, we administered resistant starch from maize to mice from the same four vendors following the same experimental procedure (see **Methods, Fig. 1B**). Compared to inulin, resistant starch stimulated milder changes in the bacterial load (**Fig. 6A**), gut microbiota composition (**Fig. 6B**), and SCFAs production (**Fig. 6C**). Under resistant starch intervention, we identified *Faecalibaculum* and *Muribaculaceae* as “primary degraders” (**Fig. 6D,** see **Table S1** for genetic evidence from literature) and 25 additional bacterial taxa as “generic responders” (**Table S3**). The dynamics of primary degraders were qualitatively similar between inulin and resistant starch interventions (**Fig. 6D**): *Muribaculaceae* increased rapidly and reached a plateau (except for Shanghai mice), while *Faecalibaculum* declined sharply after the initial burst. The gLV-based inference suggested that the observed dynamics was driven by mutual inhibition between the two primary degraders (**Fig. 6E**).

We found that bacterial load (**Fig. S14A**) and the three major SCFAs (**Fig. S14B**) exihibited baseline-dependent dynamical responses to resistant starch intervention. For example, the weak response in bacterial load and SCFA production of Shanghai mice (**Fig. 6A**) could be explained by the low abundance of *Muribaculaceae* in the baseline community (**Fig. 6F**, highlighted in red box frame). In addition, there was substantial growth of generic responders in Hunan mice (**Fig. 6G**), although the dominant bacterial taxa in this eco-group were different from the taxa identified in inulin intervention (**Table S3**).

Finally, we found that the baseline abundance of *Muribaculaceae* was separately correlated with bacterial absolute abundance and prioprionate level (**Fig. 6H**, left panel), supporting the hypothesis that *Muribaculaceae* may serve as both a primary degrader and a propionate producer. As a result, there was a weak but statistically significant positive association between bacterial load and propionate concentration (**Fig. 6H**, right panel; P=0.002). Similar to our findings from the inulin intervention group, Random Forest models based on gut microbiota compostion had low or no predictive power for SCFA concentration in the resistant starch intervention group (**Fig. 6I**). Collectively, our major findings were qualitatively consistent between inulin and resistant starch interventions, suggesting that the dynamical responses of gut microbiota to fiber-based perturbation follow universal ecological principles.

**Discussion**

Our study characterizes the ecosystem response of the gut microbiota in response to fibers, and emphasizes that ecological interactions play a key role in the personalized impact of dietary changes. By integrating gLV model with Bayesian regression, we inferred a competitive network of fiber degraders as keystone bacteria that mediate the response of gut microbiota to inulin and resistant starch intervention. gLV-based ecological inference from gut microbiome time series data has yielded mechanistic insights into the stability of probiotic community under dietary perturbation [42], colonization resistance of pathogenic *Clostridioides difficile* [43], and community assembly dynamics within preterm infant gut [22]. However, Bayesian regression using a Markov-Chain-Monte-Carlo sampler is computationally expensive, posing a limitation on the maximum number of taxa included in the gLV model (20 used in this study). Taxonomic grouping as a coarse-grained approach is thus necessary to represent the majority of microbiota composition with a limited number of taxa. As opposed to human studies, grouping sequences at a fixed taxonomic rank (e.g., species or genus level) is inappropriate in mice as too many seuqences could only be classified up to the family level (e.g., *Muribaculaceae*). In this study, we chose to group 16S sequences by the lowest classified taxonomic level, which generated sequence clusters at heterogeneous taxonomic ranks—some at species level while others at family level. Other approaches such as OTU-picking would also apply and we showed that the inference is insensitive to the choice of grouping method (**Fig. S7**).

Our findings correspond to and advance previous literature at multiple systems level understanding of the diet-induced changes in gut microbiome. The small number of fiber degraders (five for inulin and two for resistant starch) suggested that fiber-induced bacterial shifts are very selective and occur to a restricted number of taxa. Notably, not all well known fiber-degrading bacteria such as the taxa related to the genus *Bifidobacterium* were able to expand in the mouse gut on both fibers (**Fig. S15**), indicating that fiber-induced bacterial enrichement is context-dependent and cannot be simply predicted from their *in vitro* growth. Our results further revelaed that fiber-induced dynamics are largely driven by competitions and *Muribaculaceae* outcompetes other degraders in consuming both fibers. Since the family *Muribaculaceae* was specific to the mouse gut [44], it might have been adapted to the murine gut with higher fitness in utilizing fiber. Finally, we offered new insights into the fiber-induced shift in propionate production by linking its inter-individual variability with baseline abundance of *Muribaculaceae*. We propose that bacteria from this family can also produce propionate, thus correlating bacterial load and propionate concentration in the gut during dietary fiber intake.

Diet-induced changes in SCFAs are often transient and vanish shortly after cessation of dietary intervention [9, 45-47]. Our study is consistent with this result, by showing that SCFA concentrations cannot be maintained at its peak and drop by 35%-40% even under continuous inulin intake until four weeks. The transient responses in SCFAs were also observed in colorectal cancer patients [48] and type 2 diabetes mellitus patients [6]; however, it is unknown whether the reduced SCFAs in these human subjects are resulted from lower dietary fiber intake. 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. The *in vivo* SCFAs dynamics is jointly determined by multiple metabolic processes, where the two major ones are microbial production and host absorption. In healthy individuals, 90%-95% SCFAs produced in the colonic lumen are absorbed by the gut mucosa [49]. While many studies used fecal SCFAs concentrations as proxy of their luminal levels, neither of both represents the rate of production or absorption so the declined phase of SCFAs in our study may be explained by reduced production rate, increased absorption rate or both. Due to the difficulty of measuring fluxes *in vivo*, mathematical models that take both processes into accounts show great premise in the estimation of their flux rates from SCFAs concentrations [50].

Understanding the basis for the association between an observable quantity and the baseline microbiota composition is a critical step in individualized dietary fiber intervention. In previous studies, the statistical test of the association was mostly performed between the pre-to-post changes in the quantity during intervention and the relative abundance of all or selected taxa in the baseline samples. There are two main potential caveats to this conventional approach. First, the significance of association may vary depending on the study endpoint used to calculate pre-to-post changes. In our experiments, the changes in propionate concentration from their baseline levels differ significantly among the four vendors at day 5 but not at day 31 (**Fig. S16**). Second, due to the lack of control group data to assess the intervention effects, pre-to-post changes that are supposedly to capture fiber-indcued effects may be entirely attributable to random variations within each individual [51]. In constrast, our analysis avoids these two caveats by incorporating longtitudinal and comparator arm data. Additionally, the use of dimensionality reduction in our approach further benefits data visualization of inter-vendor variations in gut microbiota composition (**Fig. 3**).

Characterizing the dynamics of gut microbiota and their inter-individual variability with multi-omics data is an important priority for microbiome research to further understanding of diet-induced responses [52]. Such studies hold great promise to improve human health and treat gut microbiome-associated disease via microbiome engineering. A key question in microbiome engineering with prebiotics is whether and to what extent can we enrich the gut levels of beneficial bacteria using prebiotic compounds. Microbiome engineering, as with other engineering disciplines, requires computational tools to aid the design process. Predicting bacterial responses to interventions in the human gut is nontrivial: previous studies have repeatedly shown that bacteria able to consume a fiber supplement *in vitro* may not be selectively enriched *in vivo*, suggesting that dietary response of a gut bacterial taxa depends on its ecological traits, i.e., its ability to compete, occupy and defend for the newly opened niche (#citation needed#). By inventing a new application of gLV with uncertainty assessment to infer primary fiber degraders and associated interaction network, we provided a generalizable computational approach that facilitates ecological characterization of gut microbial community from longitudinal data. We foresee that applications of ecological modeling in human cohorts with dense longitudinal sampling will provide important insights for predictable dietary responses and personalized nutrition.

A screenshot of a computer

Description automatically generated with low confidence

**Figure 1. Longitudinal profiling of murine gut microbiota and metabolites to study the baselin-dependent dynamics in response to 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 a 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 four 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). **D**. Top four panels: presence (white blocks indicate absence) and abundance (colored blocks) of bacterial taxa in the baseline samples. Bottom panel: the prevalence score of a taxon across mice (defined as the fraction of all mice that contains this taxon) or vendors (defined as the fraction of vendors whose mice all contain this taxon).

A picture containing electronics, light, computer, dark

Description automatically generated

**Figure 2.** **Inulin-induced temporal shifts in murine gut microbiome and short-chain fatty acids (SCFAs) metabolism. A.** Bacterial load. **B.** Alpha diversity of gut microbiota composition. **C.** Relative abundance of bacterial genera shown in stacked band plot. Un.: unclassified. **D.** Shifts in gut microbiota composition represented by PCoA (principal coordinate analysis) plot. Initial and final states represent the microbiota compositions at day 0 and day 31 respectively. **E.** Fecal concentration of acetate, proprionate and butyrate. Beijing, Guangdong, Hunan, Shanghai are four different mice vendors. Lines (panels A,B), dots (panels A,B,D), and stacked bands (panels C, E) represent the mean values over mouse replicates from the same vendor (n=4 for Hunan and Guangdong, n=5 for Beijing and Shanghai). Shading areas (panels A,B) and error bars (panel D) represent standard error of the mean.

A picture containing indoor, dark

Description automatically generated

**Figure 3. Quantifying the statistical significance of baseline-dependent dynamical response. A.** A schematic diagram of our statistical framework to test for the significance of baseline-dependent response. The framework involves three steps: (1) projecting all time series from both intervention and control groups onto the same 2-dimensional space, (2) normalizing all data points by recentering the mean of control group to the origin, and (3) performing two separate statistical tests using the projected data to quantify the significance of “responsiveness” (*Pr*) and “baseline-dependence” (*Pb*). Abbreviations: Significant (Sig.); dependence (dep.). For each baseline (Beijing, Guangdong, Hunan, Shanghai), an arrow was drawn from the eclipse center of the baseline under cellulose intervention (standardized to the origin) to that under the inulin intervention. **B**,**C**. Reduced 2-dimensional representation of the inulin-induced responses in bacterial load (B) and three major SCFAs (C). In all panels, each symbol represents a mouse (dots: cellulose group, crosses: inulin group) and all mice data from the same vendor under the same intervention (inulin or cellulose) was used to fit an eclipse (ellipse’s radius was determined by 2 standard deviations).

**A screenshot of a computer game

Description automatically generated with low confidence**

**Figure 4. Inulin** **responders** **shape the dynamical response of murine gut microbiota. A.** Generalized Lotka-Volterra (gLV) model combined with Bayesian statistics to infer inulin degraders and pairwise 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 positive. **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 according 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 group dynamics of primary inulin degraders, generic responders (presented with two subgroups) and non-responders. **E**. Mean absolute baseline abundance of the seven bacterial taxa shown in the panel B. **F**. Temporal changes in the absolute abundance of the top three inulin degraders. In panels D, F, lines and dots represent the mean values across mice from the same vendor (n=4 for Hunan and Guangdong, n=5 for Beijing and Shanghai). Shading areas represent the standard error of the mean.

**A screenshot of a video game

Description automatically generated with medium confidence**

**Figure 5.** **Quantitative relationship between SCFAs and murine gut microbiota composition.** **A.** Correlation between bacterial load and propionate concentration (right big panel). We proposed that the correlation is mediated by some inulin responders which causally and simultaenoulsy affect both observations. Eight small panels to the left: Spearman’s correlations of baseline abundance of four inulin responders with the mean bacterial load (top row) or propionate concentration (bottome row) averaged across the interveion period. Dashed line: Lowess regression. Spearman correlation coefficient (ρ) and adjusted P-value are indicated in each plot. **B**. Prediction of SCFAs concentrations from gut microbiota composition using machine learning models. Two data-split strategies for testing model performance were designed: 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. **C**.Performances of Random Forest regression models on the training and test datasets.

**Graphical user interface

Description automatically generated**

**Figure 6. Resistant starch-induced dynamical response in murine gut microbiota. A-C.** Dynamical responses of bacterial load (A), gut microbiota composition (B), and SCFAs concentration (C) following resistant starch intervention. **D**.Dynamics of two putative resistant starch degraders. represents the growth impact of resistant starch and its posterior distributions are shown for each degrader. CI: credible interval. **E**. Ecological interactions between the two degraders. The arrow thickness is proportional to the posterior mean of the corresponding interaction coefficient. **F**. Mean baseline abundances of the two degraders. **G**. Ecological group dynamics of primary degraders and generic responders of resistant starch as well as the non-responders. **H**. Correlations among baseline abundance of *Muribaculaceae*, bacterial load and propionate concentration. The time-averaged responses were calculated by dividing the area under the curve by the duration of observation. Dashed lines: Lowess regression. **I**. Prediction of SCFAs concentration from gut microbiota compsotion using a random forest regression model. “Interpolation” and “extrapolation” are two strategies of splitting all data into the training and test sets (the same as **Fig. 5B**). Lines (panels A,D,G) or height of stacked bands represent mean values across mice from the same vendor (n=5 for all vendors). Shading areas (panels A,D,G) represent the standard error of the mean.

**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 S4**). 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 previous three days per cage were measured. This study was approved by the Institutional Animal Care and Use Committee 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 [53]. 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 previously described [54]. 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 [55]. 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 [56], 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. Sequencing was performed in a single run on NovaSeq 6000 (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 on NovaSeq 6000 (Illumina, USA) by Guangdong Magigene Biotechnology Co.,Ltd. (Guangzhou, China).

**Bioinformatics analysis**. The 16S rRNA sequencing reads were analyzed by QIIME 2 (version 2020.2) [57]. 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 amplicon sequence variants (ASVs) in the blank control or the negative correlation with DNA concentration [58]. 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 [59]. 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. S17**), indicated sufficient sequencing depth for majority of the analyzed samples. Estimated alpha diversity metrics using q2-diversity. Beta diversity metrics (Aitchison distance) and biplot were generated using DEICODE (robust Aitchison PCA, RPCA) [60]. 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 ASVs, the q2-feature-classifier basing on the classify-sklearn naïve Bayes taxonomy classifier was used with the SILVA (v.138) as reference database. Unless specified (**Fig. 2C**, **Fig.6B**, **Fig. S1A**), ASVs were grouped to the lowest classified taxonomy level for all data modeling and analysis. These taxonomic groups are labelled with prefix “s\_” (species level), “g\_” (genus level), “f\_” (family level), “o\_” (order level), “c\_” (class level), “p\_” (phylum level), and “k\_” (kingdom level) to indicate the taxonomic rank where grouping was operated. For example, “g\_Bacteroides” clusters all sequences that are classified as Bacteroides at the genus level but unclassified at the species level.

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 [61]. 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 [62]. The gene abundance was analyzed and calculated as previously described with modifications [63]. Putative genes were then predicted on contigs longer than 200 base pairs using Prodigal under metagenome mode (-p meta) [64]. A non-redundant gene catalogue was constructed with CD-HIT using the parameters “-c 0.95 –aS 0.9” [65]. 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) [66]. For all the predicted genes, CAZymes were annotated using hmmsearch against the dbCAN2 database V9 (e value <1 × 10−10; coverage >0.3) [67]. 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 [68]. The binning results were refined based on the bin quality assessment (completeness >75, and contamination <15) of different binners from CheckM [69]. Taxonomic classification of each bin was determined by GTDB-tk [70], and subjected to prediction of polysaccharide utilization loci (PULz) using pipeline PULpy [71].







**Significance test of baseline-dependent responses.** Sequential non-negative matrix factorization [33] 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 and (2) two-dimensional data can be easily visualized. Fhe reduced representation of the intervention group and control group , () 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 statistical significance test of: 1) the responsiveness (i.e., whether time series in the intervention group differs from that in the control group regardless of vendor), and 2) the 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 statistical module “statsmodels” in python). 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 in the intervention group (“baseline dependence”, Pb). If both P-valus are smaller than 0.05, we determined that a quantity has a baseline-dependent response. For all significance tests that require multiple test correction, the Benjamini-Hochberg procedure [72] was used for controlling the false discovery rate in multiple test correction.

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

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

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 dietary 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 [73]. For each mice (), Eq. (1) can be transformed into a matrix form that incorporates all discrete time points of measurements (, )

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

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. (3) |

The linear regression as described in Eq. (4) (for brevity ) can be further transformed into a Bayesian regression where and represent normal distribution and standard deviation respectively.

Since gLV models the absolute abundance of bacterial taxa, we multiplied the bacterial load by their relative abundance to calculate absolute abundance. The time-series data from all mice were simultaneously fed into the gLV model based on the premise that ecological forces driving microbiome dynamics are largely host-independent and universal [74]. We used uninformative priors for all gLV parameters and Stan program [75] 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 taxa with the highest time-averaged absolute abundances in the inulin (or resistant starch) and cellulose group.

**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. Two data-split approaches were implemented. For the “interpolation” approach, the mice from the same vendor were first alphabetically labeled as A-D (for Hunan and Guangdong) or A-E (for Beijing and Shanghai). Then the four mice with the same label (one per vendor) were chosen to constitute the test set and the remaining mice were mixed together to form the training set. For the “extrapolation” approach, all mice from a specific vendor were chosen as the test set while the training set includes all mice from the other three vendors. To train each model, five hyperparameters were tuned using 5-fold cross validation within the training set and R2 as the scoring metric (GridSearchCV function of the scikit-learn library): 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.

**Data availability**

Sequencing data: desposit raw data to SRA

SCFA and other data: supplementary tables

**Code availability**

The customized python scripts for data modeling and analysis are available in the Github repository: https://github.com/liaochen1988/Source\_code\_for\_SCFA\_dynamics\_project

**References:**

[1]. Tolhurst, G., et al., Short-chain fatty acids stimulate glucagon-like peptide-1 secretion via the G-protein-coupled receptor FFAR2. Diabetes, 2012. 61(2): p. 364-71.

[2]. Vinolo, M.A.R., et al., Regulation of Inflammation by Short Chain Fatty Acids. Nutrients, 2011. 3(12): p. 858-876.

[3]. Litvak, Y., M.X. Byndloss and A.J. Bäumler, Colonocyte metabolism shapes the gut microbiota. Science, 2018. 362(6418): p. eaat9076.

[4]. 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.

[5]. Parada Venegas, D., et al., Short Chain Fatty Acids (SCFAs)-Mediated Gut Epithelial and Immune Regulation and Its Relevance for Inflammatory Bowel Diseases. Frontiers in Immunology, 2019. 10.

[6]. Zhao, L., et al., Gut bacteria selectively promoted by dietary fibers alleviate type 2 diabetes. Science, 2018. 359(6380): p. 1151-1156.

[7]. Sitkin, S., T. Vakhitov and J. Pokrotnieks, How to Increase the Butyrate-producing Capacity of the Gut Microbiome: Do IBD Patients Really Need Butyrate Replacement and Butyrogenic Therapy? Journal of Crohn's and Colitis, 2018. 12(7): p. 881-882.

[8]. 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.

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

[10]. Singh, V., et al., Microbiota fermentation-NLRP3 axis shapes the impact of dietary fibres on intestinal inflammation. Gut, 2019: p. gutjnl-2018-316250.

[11]. 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.

[12]. 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).

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

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

[15]. 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).

[16]. 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.

[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]. Kovatcheva-Datchary, P., et al., Dietary Fiber-Induced Improvement in Glucose Metabolism Is Associated with Increased Abundance of Prevotella. Cell Metabolism, 2015. 22(6): p. 971-982.

[19]. Davis, L.M.G., et al., Barcoded pyrosequencing reveals that consumption of galactooligosaccharides results in a highly specific bifidogenic response in humans. PloS one, 2011. 6(9): p. e25200.

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

[21]. 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.

[22]. Rao, C., et al., Multi-kingdom ecological drivers of microbiota assembly in preterm infants. Nature, 2021. 591(7851): p. 633-638.

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

[24]. 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.

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

[26]. 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.

[27]. 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.

[28]. 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).

[29]. 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.

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

[31]. 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).

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

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

[34]. 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.

[35]. 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.

[36]. 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.

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

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

[39]. Patnode, M.L., et al., Interspecies Competition Impacts Targeted Manipulation of Human Gut Bacteria by Fiber-Derived Glycans. Cell, 2019. 179(1): p. 59-73.e13.

[40]. Salonen, A., et al., Impact of diet and individual variation on intestinal microbiota composition and fermentation products in obese men. ISME J, 2014. 8(11): p. 2218-30.

[41]. 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.

[42]. Bucci, V., et al., MDSINE: Microbial Dynamical Systems INference Engine for microbiome time-series analyses. Genome Biology, 2016. 17(1).

[43]. Buffie, C.G., et al., Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile. Nature, 2015. 517(7533): p. 205-8.

[44]. 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).

[45]. 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).

[46]. 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.

[47]. 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.

[48]. 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.

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

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

[51]. 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.

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

[53]. 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.

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

[55]. 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.

[56]. 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.

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

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

[59]. 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.

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

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

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

[63]. 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.

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

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

[66]. 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).

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

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

[69]. 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.

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

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

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

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

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

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