**Deciphering the baseline-dependent ecological dynamics of murine gut microbiome in response to dietary fiber interventions**

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**

Intake of dietary fibers selectively modulates the composition and metabolism of gut microbiome and the effects are highly individualized. The individuality in response to dietary fiber interventions has been associated with the baseline gut microbiome, however, in-depth understanding of the ecological drivers in microbiota remodeling is still lacking. In this study, we fed fermentable dietary fibers to isogenic mice with distinct baselines of gut microbiota and profiled the long-term dynamics of gut microbiome and short-chain fatty acids (SCFAs). Under inulin intervention, the murine gut microbiota (bacterial absolute abundance, community composition) and SCFAs showed a two-phase response, i.e. rapid stimulation in the short-term phase and gradual stabilization to new equilibria in the long-term phase, and the dynamics were baseline-dependent. Based on absolute abundance quantification and ecological modeling, we inferred a group of bacterial taxa as key responders, whose growth was selectively promoted by inulin. We found that the baseline-dependent dynamics of bacterial load and community composition were largely determined by the baseline abundance of key inulin responsders and their interspecies competition. The production of SCFAs, such as propionate, can be directly associated to the abundance of key responders, yet quantitative predictions of metabolite levels from bacterial community composition were impeded by inter-individual variations. Finally, we demonstrated that our methods and findings were generalizable to study the dynamical response to resistant starch intervention. Overall, our study emphasizes the importance of ecological modeling for understanding the dynamical responses of gut microbiome 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. One of the major anaerobic fermentation products of gut bacteria in the large intestine is short-chain fatty acids (SCFAs). SCFAs, mainly consisting of 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]. Impaired SCFAs production has been associated with gut microbiota dysbiosis and a wide range of dieseases [2][3][4]. Dietary fibers have been proposed as “prebiotics” that can selectively enrich beneficial gut bacteria and administered as a therapeutic strategy to modulate and restore intestinal gut microbiota and SCFA levels in clinical trials. It has been observed that dietary fibers can rapidly alter gut microbiota, including both the community composition and the absolute abundance.

However, several studies in healthy adults have collectively revealed that the ability of dietary fibers to induce SCFAs production 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 the baseline gut microbiota is a critical factor for explaning the observed heterogeneity at multiple levels [12, 13]. For example, the inter-individual variation in response to dietary fiber interventions is related to the differential abundance of key fiber-degrading bacteria. The impaired ability of healthy human donors’ fecal samples to ferment resistant starch can be restored by *in vitro* co-incubation with *R. brommi,* a well-known degrader of resistant starch. The inter-individual variations in the composition and metabolites of gut microbiome 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 selects for microbial taxa with specific nutrient utilization functions and subsequently leads to remodelling of the entire gut microbial community. The shift in microbial community under the stimulation of dietary fibers can be abstracted by the dynamics of a baseline-specific ecological network (**Fig. 1A**). Consumption of dietary fibers selectively enriches a fairly limited number of bacteria (i.e. primary degraders) which rapidly expand and dominate the gut microbiota after substantial induction. For example, the relative abundance of *Bifidobacterium* species consistently increases in healthy people following the intake of dietary fibers. Production of 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 [5, 6, 7, 8]. Despite emerging interests microbiota ecology and targeted modulation, a system-level, quantitative understanding of the ecological dyanmics of gut microbiome under dietary interventions and the dependence on the baseline community composition is still lacking.

In this study, we combined longitudinal profiling of murine gut microbiota and ecological modelling to study the key factors underlying the baseline-dependent dynamical response to dietary fibers. We developed computational approaches that employed time series data to analyze the heterogeneity in dietary responses and study its relationships with baseline composition. The stimulated growth and interspecific competitions of key responders constitute the major shifts in the gut microbiota composition due to their ability of fiber degradation and growth advantage over other community members. Moreover, we showed that the heterogeous responses in bacterial absolute abundance and SCFA production are linked to the baseline abundance of a few putative fiber-degrading bacteria. Our 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**

**Heterogeneity in the baseline murine gut microbiome**

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 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 [19]. 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 known to be degraded by gut bacteria in the cecum and colon [20, 21] and used to stimulate the production of SCFAs [10, 22].

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, Fig. 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. 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-dependet dynamics of murine gut microbiome 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 taxa, 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 have profound impacts on the ecology and function of gut microbiota [29, 30]. In addition, we found the tendency of gut microbiota 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. To quantify the rate of stabilization, we fit the time series of community diversity to a harmonic oscillator model to calculate the damping ratio (i.e. a dimensionless measure of whether and how dynamical systems approach new steady states upon perturbations) (see **Methods, Table S1**). Although the damping ratio varied among individual mice, the dynamical responses of microbiota diversity in ~60% of our mice were critical damping, i.e. non-oscillatory convergence towards the new equilibria. The critical damping behavior was also found in the recovery of human gut microbiome after transient antibiotic exposure [31].

We observed a strong stimulating effects of inulin on the metabolism of three major SCFAs and valerate (**Fig. 2E, 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 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 if the gut microbiota exihibit baseline-dependent responses, we developed a novel approach that 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 to capture representative temporal trendsand for data visualization (**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*). Using this approach, we confirmed that the dynamical responses of bacterial load (**Fig. 3B**), propionate and butyrate (**Fig. 3C**) were baseline-dependent.

**Primary degraders of inulin are ecological drivers of the baseline-dependent dynamical response**

We used the generalized Lotka–Volterra (gLV) model to infer key ecological drivers of gut microbiota response in dietary fiber intervention (**Fig. 4A, 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. In our gLV-based framework, any bacteria taxa with a significantly positive “fiber response” is considered a putative “primary degrader” of inulin. We identified five bacterial taxa as primary degraders, including *Bacteroides acidifaciens* (species), *unclassified Muribaculaceae* (family), *unclassified Faecalibaculum* (genus), *unclassified Parasutterella* (genus), and *unclassified Bacteroides* (genus) (**Fig. 4B**). For four of these bacterial taxa, we found genetic evidence and/or *in vitro* experiments from literature to support their functional roles in inulin degradation (**Table S3**). 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 S4**). Furthremore, we analyzed the data from an independent study [33], 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. S11**). 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.

Alternatively, we used the statistical test of responsiveness (**Figure 3A**) to identify “generic responders” (*Pr* <0.05, after multiple test correction), we found a total of 37 bacterial taxa with significant dynamical responses (**Table S2**). Some of the responders are known for inulin degradation, such as species of the genus *Bacteroides* and *Bifidobacterium*. Two bacterial species whose relative abundances significantly increased after day 5 (**Fig. S7, S8**), *Akkermancia muciniphila* and *Bacteroides uniformis,* were inferred as “generic responders”. 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 some responders could be bacterial taxa that benefit from the primary degraders. Indeed, inference of the gLV model did not identify *A. muciniphila* as “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 have breen demonstrated as a useful concept to understand microbial ecology [39, 40]. We divided the entire gut microbiota into three eco-groups: (1) 5 primary degraders of inulin; (2) 37 generic responders to inulin intervention (excluding the primary degraders); (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 baseline composition of key bacterial taxa (**Fig. 4D-E**). For example, the abundances of *A. municiphila* and *B. uniformis* increased dramatically in Hunan mice (**Fig. S8**), 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 *un. Muribaculaceae* in Shanghai mice (violet box frames in **Fig. 4E**) may explain the sluggish responses in bacterial absolute abundance and SCFAs (**Fig. S8, Fig. S5**).

Our gLV-based inference suggested strong competition among primary degraders, where *un.* *Muribaculaceae* inhibits the growth of *B. acidifaciens* and *un.* *Facaelibaculum* (**Fig. 4C**). Indeed, *B. acidifaciens* and *un.* *Facelibaculum* showed transient dynamics with a quick rise and drop in their absolute abundances, while the abundance of *un*. *Muribaculaceae* increased steadily and remained high during the entire period of observations (**Fig. 4F, S8**). Our results are consistent with previous studies by Patnode *et al.* that 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). Taken together, we demonstrate that primary degraders and their interspecies 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 dynamical response of SCFAs to inulin intervention varied substantially across different baselines (**Fig. S5**). We noted that Shanghai mice produced the lowest level of propionate; these mice also showed the lowest response in bacterial load, due to lack of primary degraders of inulin in the baseline microbiota (**Fig. 4D**). Thus we hypothesized that the primary degraders of inulin may directly contribute to propionate production. Among the key bacterial taxa involved in inulin response (**Fig. 4B**), we found that the baseline abundances of *B. acidifaciens*, *un. Muribaculaceae*, *A. municiphila*, *B. uniformis* were positively correlated the propionate concentration (**Fig. 5A** left panel). Indeed, *un. Muribaculaceae*, *A. municiphila* and *B. uniformis* have been previously found to produce propionate *in vitro* and/or *in vivo* (**Table S3**). As a result, there was a strong positive association between bacterial load and propionate concentration (**Fig. 5A** right panel, P<0.001), consistent with previous studies. In contrast, the association between bacterial load and other SCFAs (butyrate: P=0.190, acetate: P=0.047) was not significant.

Given the clear associations between gut microbiota and metabolism, we wonder if it is possible to quantitatively predict SCFAs from bacterial community composition. 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. S9A**), 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. S9B**), 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. S10A**), models (**Fig. S10B**) or adding weights to training samples (**Fig. S10C**). 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. In support of previous analysis of human data sets, our findings suggest that the substantial inter-individual variation of gut microbiome could impede the predictive power of machine learning models,

**Analysis of resistant starch-induced dynamical response in murine gut microbiota**

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 (**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, 27 bacterial taxa were identified as “generic responders” (**Table S2**), among which *un. Faecalibaculum* and *un. Muribaculaceae* were inferred as “primary degraders” by gLV model (**Fig. 6D,** see **Table S3** for genetic evidence from literature). The dynamics of primary degraders were qualitatively similar between inulin and resistant starch interventions (**Fig. 6D**): *un. Muribaculaceae* increased rapidly and reached a plateau (except for Shanghai mice), while *un. 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. S12A**) and the three major SCFAs (**Fig. S12B**) 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 *un. 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 S2**).

Finally, we found that the baseline abundance of *un. Muribaculaceae* was positively correlated with both bacterial absolute abundance and prioprionate (**Fig. 6H**, left panel), supporting the hypothesis that *un. Muribaculaceae* was 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 the findings of the inulin intervention group, we found that 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.

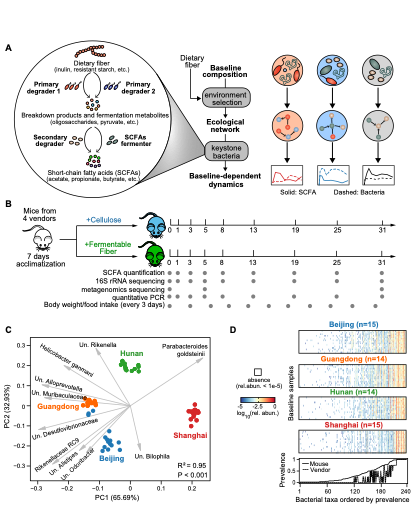
**Discussion**

Our study emphasizes that analysis of ecological dynamics is paramount to understanding the dietary responses of gut microbiota. 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, colonization resistance of pathogenic *Clostridioides difficile*, and community assembly dynamics within preterm infant gut. Besides evidences supporting the fiber-degrading function of these putative degraders, our findings correspond to and advance previous literature at multiple systems level understanding of the effects of dietrary changes. First, 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. Second, the absolute abudance of many fiber-degrading bacteria such as taxa related to the genus *Bifidobacterium*, failed to expand in the mouse gut on both fibers (Fig. S13), indicating that fiber-induced bacterial enrichement cannot be simply predicted from their *in vitro* growth. Third, our results revelaed that fiber-induced dynamics are largely driven by competitions and *un. Muribaculaceae* outcompetes other degraders in consuming both fibers. Since the family *Muribaculaceae* was specific to the mouse gut [74], it might have been adapted to the murine gut with higher fitness in utilizing fiber. Fourth, we offered new insights into the fiber-induced shift in propionate production by linking its inter-individual variability with baseline abundance of *un. 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.

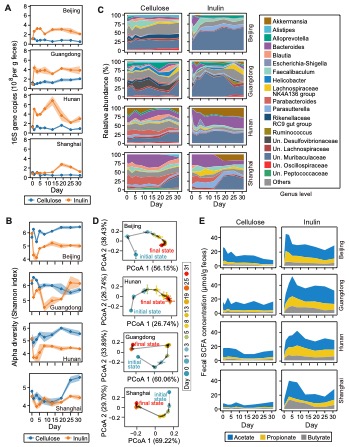
Diet-induced changes in SCFAs are often transient and vanish shortly after cessation of dietary intervention [52-55]. Our experiments add to the literature by further showing that SCFA concentrations cannot be maintained at its peak and drop by 35%-40% even under continuous inulin intake until four weeks. 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. 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 [58]. 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.

Quantifying the association of an observable quantity with baseline microbiota composition is a critical step in individualized response analysis of 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. S14). 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 [43]. 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 visualization of inter-vendor variations in gut microbiota composition (Fig. 3B,C).

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 [51]. 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 a beneficial bacterial population can be selectively induced by a prebiotic compound. Microbiome engineering, as with other engineering disciplines, requires computational tools to aid the design process. Prediction of bacterial growth in 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. 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 facilitate 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.



**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). Un.: unclassified/uncultured. **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).



**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. **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=5?). Shading areas (panels A,B) and error bars (panel D) represent standard error of the mean. Un.: unclassified/uncultured.

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 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 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).

**Graphical user interface

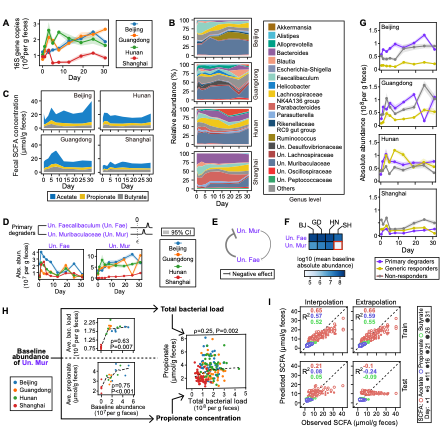
Description automatically generated**

**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 (Beijing, Guangdong, Hunan, Shanghai are four different mice vendors) and shading areas represent the standard error of the mean. Un: unclassified/uncultured.

**A screenshot of a video game

Description automatically generated with medium confidence**

**Figure 5.** **The relation 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. Beijing, Guangdong, Hunan, Shanghai are four different mice vendors. 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.

****

**Figure 6. Analysis of 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 of resistant starch, generic responders (resistant starch responders excluding the two degraders), and non-responders. **H**. Correlations among baseline abundance of *un. 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 (see Fig. 5B for details). In all panels, Beijing (BJ), Guangdong (GD), Hunan (HN), Shanghai (SH) are four mice vendors. Lines (panels A,D,G) or height of stacked bands represent mean values across mice from the same vendor and shading areas (panels A,D,G) represent the standard error of the mean. Un.: unclassified/uncultured.

**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 S5). 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 [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 previously described [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. 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 XXX.

**Bioinformatics analysis**. The 16S rRNA sequencing reads were analyzed by QIIME 2 (version 2020.2) [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. S16), 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. Unless specified, ASVs are grouped to the lowest classified taxonomy level for further analysis.

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].

**The phenomenological model of a damped harmonic oscillator.** 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 (indicate the software used) with constraints . The best-fit values of and were then used to calculate the damping ratio .

The damped harmonic oscillator model fitted the data reasonably well with mean R2 of 73% (Table S1). The damping ratios were all equal or greater than 1, meaning that the dynamical responses were critically damped or overdamped (i.e.no oscillations).

**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 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 (indicate the software used). 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 [96] 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. (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 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 [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. (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 [34]. 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.

**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 to 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.

**Data availability**

Sequencing data: desposit raw data 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.