

ARTICLE



Ecological dynamics of the gut microbiome in response to dietary fiber

Hongbin Liu^{1,6}, Chen Liao^{2,6}, Lu Wu¹, Jinhui Tang³, Junyu Chen¹, Chaobi Lei¹, Linggang Zheng¹, Chenhong Zhang⁴, Yang-Yu Liu⁵, Joao Xavier² and Lei Dai¹✉

© The Author(s), under exclusive licence to International Society for Microbial Ecology 2022

Dietary fibers are generally thought to benefit intestinal health. Their impacts on the composition and metabolic function of the gut microbiome, however, vary greatly across individuals. 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 the gut microbiome and short-chain fatty acids (SCFAs) in isogenic mice with distinct microbiota baselines fed with the fermentable fiber inulin and resistant starch compared to the non-fermentable fiber cellulose. We found that inulin produced a generally rapid response followed by gradual stabilization to new equilibria, and those dynamics were baseline-dependent. We parameterized an ecology model from the time-series data, which revealed a group of bacteria whose growth significantly increased in response to inulin and whose baseline abundance and interspecies competition explained the baseline dependence of microbiome density and community composition dynamics. Fecal levels of SCFAs, such as propionate, were associated with the abundance of inulin responders, yet inter-individual variation of gut microbiome impeded the prediction of SCFAs by machine learning models. We showed that our methods and major findings were generalizable to dietary resistant starch. Finally, we analyzed time-series data of synthetic and natural human gut microbiome in response to dietary fiber and validated the inferred interspecies interactions *in vitro*. This study emphasizes the importance of ecological modeling to understand microbiome responses to dietary changes and the need for personalized interventions.

The ISME Journal (2022) 16:2040–2055; <https://doi.org/10.1038/s41396-022-01253-4>

INTRODUCTION

Fermentable dietary fibers, such as inulin and resistant starch, are edible carbohydrate polymers that escape digestion by host enzymes in the upper gut and are fermented by gut microbiota in the cecum and colon. The major products from the microbial fermentative activity in the large intestine are short-chain fatty acids (SCFAs)—mainly acetate, propionate, and butyrate—which have broad impacts on intestinal health and immunity [1–3]. Impaired SCFAs production has been associated with multiple diseases [4, 5]. Dietary fibers can selectively enrich beneficial gut bacteria [6] and could be administered as “prebiotic” therapies to restore intestinal gut microbiota and elevate SCFA levels [7, 8].

Previous work showed that dietary fibers can cause rapid changes in microbiota composition and biomass [9, 10]. But the ability of fibers to increase SCFA production varies across individuals [11–14]. For example, Baxter et al. showed that resistant starch was able to promote butyrate production in only 63% of participants [12]. While the individualized response can arise from a combination of factors such as host genetics and diet history, the baseline gut microbiota is also a critical factor [14, 15].

The feces of some healthy human donors fail to ferment resistant starch, which can be restored by co-incubation with *Ruminococcus bromii*, a well-known degrader of resistant starch [16]. Person-to-person variation in the bacterial and metabolic composition of gut microbiome [17] can further impact biological variables such as body mass index [18] and glucose tolerance [18, 19] of the human host.

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 complex ecological network [20] (Fig. 1A). The primary users of dietary fibers are relatively depleted in individuals on a low-fiber diet, but they can rapidly expand and dominate the gut microbiota after substantial induction [21]. 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 oligosaccharides) and fermentation metabolites (e.g., pyruvate), which can respectively benefit the secondary fiber degraders and

¹CAS Key Laboratory of Quantitative Engineering Biology, Shenzhen Institute of Synthetic Biology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China. ²Program for Computational and Systems Biology, Memorial Sloan-Kettering Cancer Center, New York, NY, USA. ³Guangdong Provincial Key Laboratory of Brain Connectome and Behavior, CAS Key Laboratory of Brain Connectome and Manipulation, the Brain Cognition and Brain Disease Institute, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China. ⁴State Key Laboratory of Microbial Metabolism, School of Life Science and Biotechnology, Shanghai Jiao Tong University, Shanghai, China. ⁵Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA. ⁶These authors contributed equally: Hongbin Liu, Chen Liao. ✉email: lei.dai@siat.ac.cn

Received: 4 September 2021 Revised: 5 May 2022 Accepted: 11 May 2022

Published online: 21 May 2022

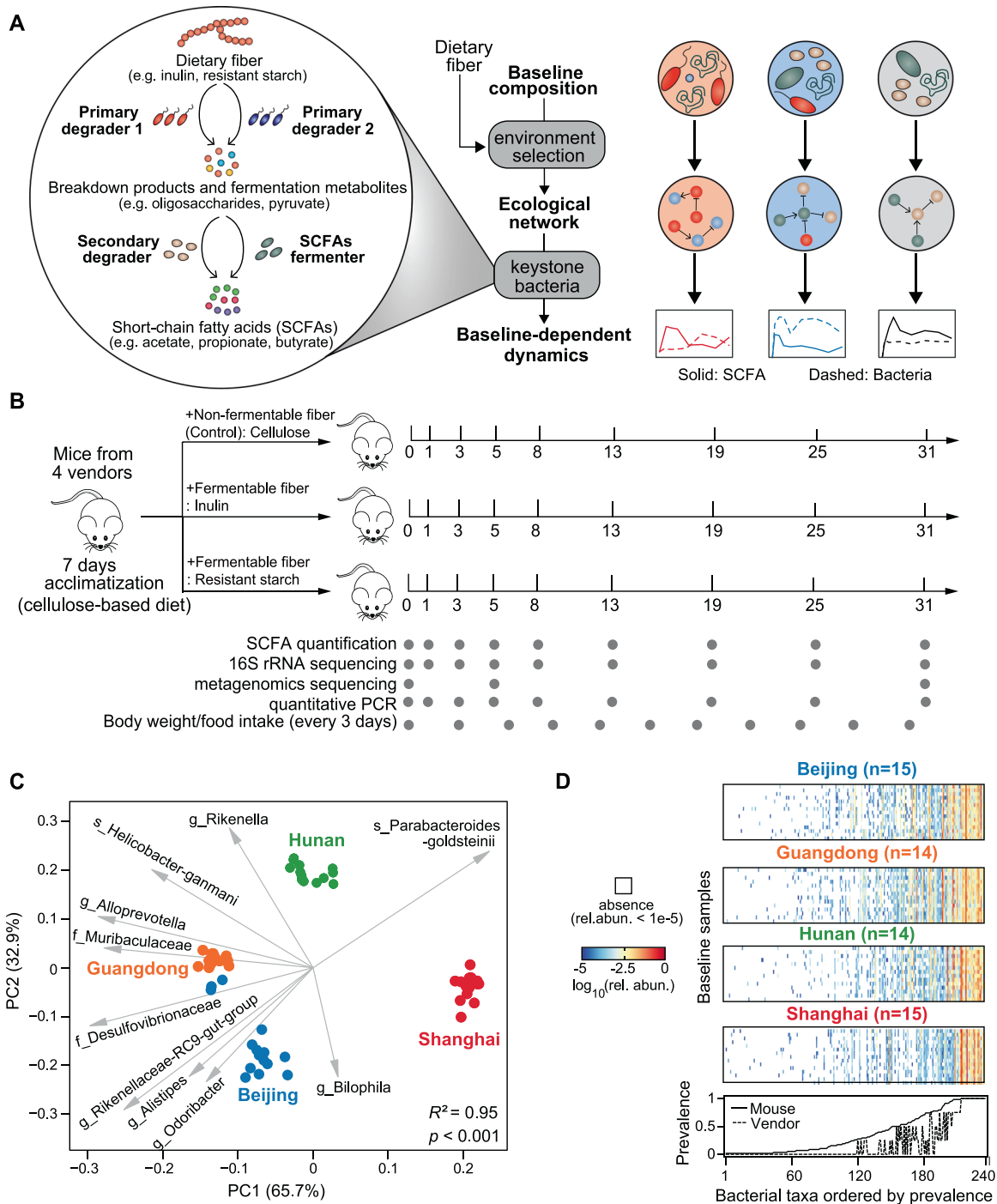


Fig. 1 Longitudinal profiling of murine gut microbiota and metabolites to study the baseline-dependent dynamics in response to dietary fiber intervention. **A** An ecological perspective of baseline-dependent dynamical responses of gut microbiota and SCFAs. Administration of dietary fibers alters the 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 drive heterogeneous responses of bacteria and SCFAs via diverse ecological interactions (e.g., resource competition 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 principal coordinate analysis (PCoA) biplot. Isogenic age- and gender-matched mice were purchased from four different vendors (Beijing, Guangdong, Hunan, and 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 contain this taxon) or vendors (defined as the fraction of vendors whose mice all contain this taxon).

SCFAs producers [22, 23]. Despite increasing interests in microbiota ecology and targeted modulation [24–27], a system level, quantitative understanding of the ecological dynamics of gut microbiome under dietary interventions and the dependence on the baseline community composition is still lacking.

In this study, we profile longitudinally the gut microbiota of mice to study the ecological basis for the baseline-dependent dynamical response to dietary fibers. We use the time-series data to infer the ecological network that explains why the microbiota fiber responses vary with their baseline composition. We find that

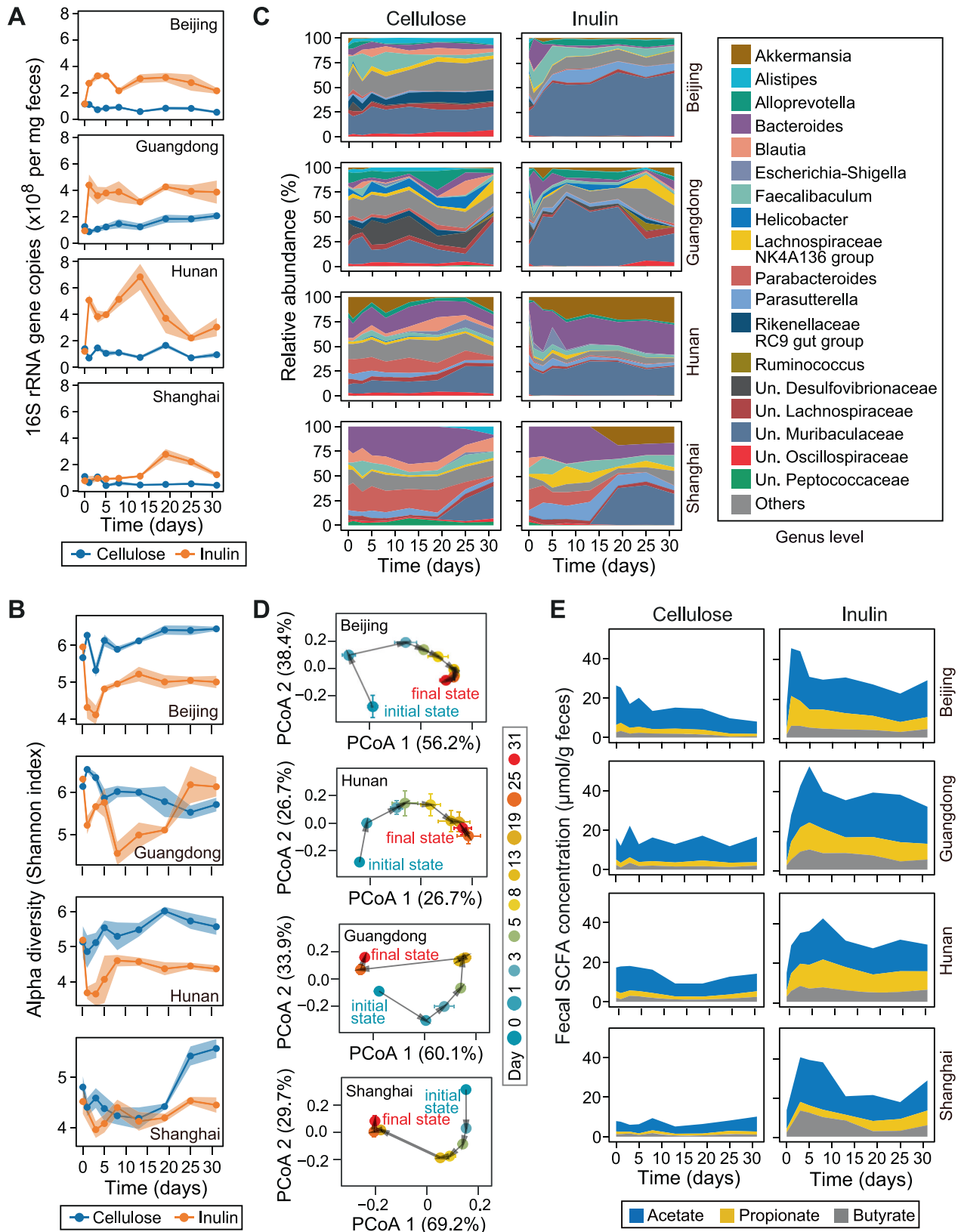


Fig. 2 Inulin-induced temporal shifts in murine gut microbiome and short-chain fatty acids (SCFAs) metabolism. **A** Bacterial load. **B** Alpha diversity of gut microbiota composition. **C** Relative abundance of bacterial genera shown in stacked band plot. Un.: unclassified. **D** Shifts in gut microbiota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

the growth rates and ecological interactions of key bacteria with known ability to benefit from fibers drive the major shifts in the gut microbiota composition. We further show that their baseline abundances predict individualized responses in bacterial density and SCFA concentration. Finally, we analyze the ecological response of synthetic and natural human gut microbiome to dietary fiber from previously published time-series data and validate the inferred interspecies interactions through in vitro experiments. 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 [28]. To ensure the distinctness of their baseline gut microbiomes, 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 (by quantitative PCR) and community composition of gut bacteria (by 16S rRNA gene amplicon sequencing and shotgun metagenomics sequencing), SCFAs concentration (by targeted metabolomics), as well as physiological changes following the intervention of fermentable dietary fibers (inulin and resistant starch from maize) and cellulose (control group) (Fig. 1B). The two dietary fibers used in this study, inulin and resistant starch from maize, can be degraded by gut bacteria in the cecum/colon [29, 30] and to stimulate the production of SCFAs [8, 12].

Consistent with previous studies [31, 32], these mice grown in vendor-specific housing and feeding conditions can be naturally divided by vendor sources into groups with distinct microbiota composition after 7 days of acclimatization. Principal coordinate analysis based on the robust Aitchison distance [33] shows that the baseline compositions of those mice can be naturally clustered by vendors (Adonis, $p < 0.001$) and are characterized by distinct bacterial taxa (Figs. 1C and S1A). For example, Shanghai mice have low relative abundances of several commensal polysaccharide-degrading bacteria such as *Muribaculaceae* and *Rikenellaceae* [34, 35]. The profound inter-vendor differences are also noticeable at the level of presence and absence of bacteria: ~65% of 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). Consistent with previous findings [36], the prevalence and abundance of bacterial taxa in these baseline samples exhibit a strong positive linear relationship on a logarithmic scale (Fig. S1B). 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 was 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 *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. Specifically, 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 (inulinase/fructanase) significantly increased after intervention (Fig. S4B). Collectively, our longitudinal profilings are consistent with previous observations that dietary fibers have profound impacts on the dynamics and function of gut microbiota [37, 38]. In addition, we found the tendency of gut microbiota to stabilize under sustained stimulation of inulin (Fig. 2D). The microbiota compositions at the end of inulin intervention (all mice sacrificed on day 31) were clearly distinct from their baselines, indicating 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 (Figs. 2E and S5). Since SCFAs are metabolites produced by colonic bacterial fermentation of inulin, we expect 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 temporal 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 on day 31 were still 2.0–3.5-fold of their 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 SCFA production may depend on the baseline microbiota. We developed a statistical analysis, which we name “NMF-PERMANOVA”, to test whether a bacterial species or metabolite responds to dietary fiber (“responsiveness”) and whether the response depends on baseline microbiota (“baseline dependence”) within the same framework (see “Methods”). Briefly, time-series data of both intervention and control groups were projected onto a 2-dimensional space by sequential non-negative matrix factorization (NMF) [39] to capture

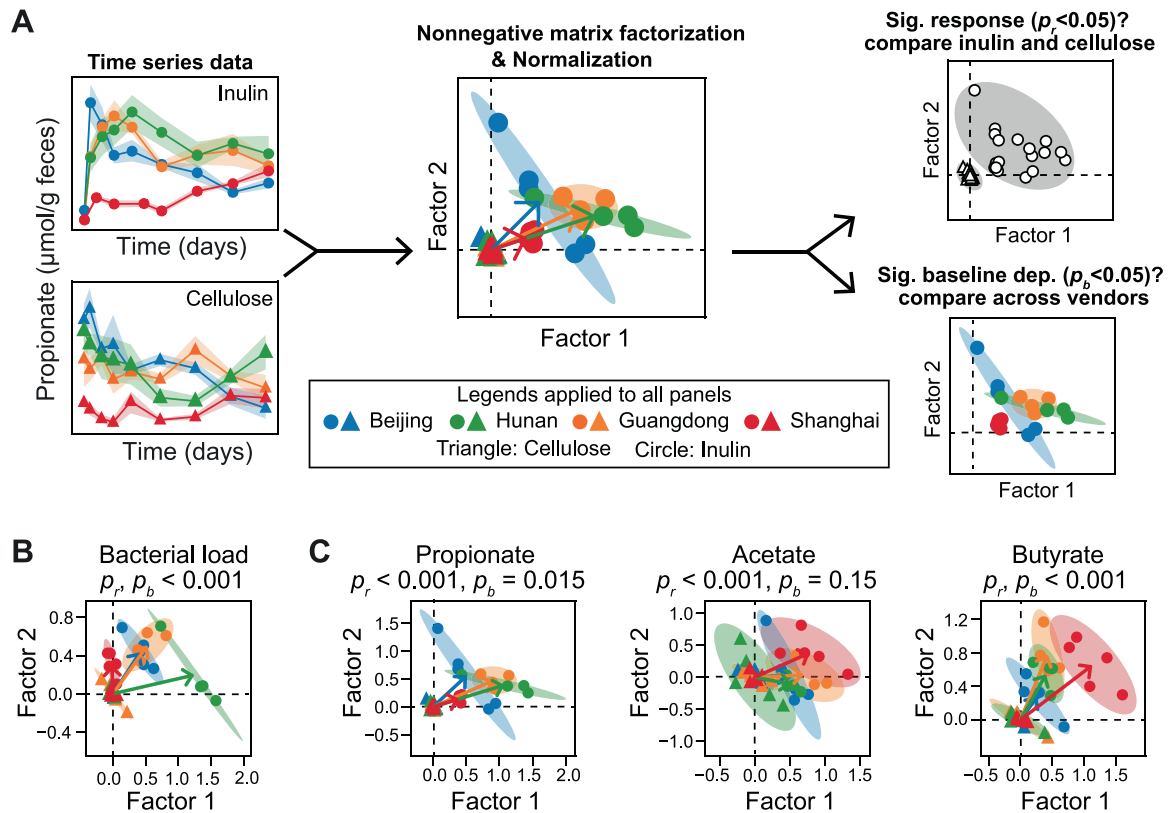


Fig. 3 Quantifying the statistical significance of baseline-dependent dynamical response. **A** A schematic diagram of our statistical framework to test for the significance of baseline-dependent response. The framework involves three steps: (1) projecting all time-series from both intervention and control 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” (p_r) and “baseline dependence” (p_b) using Permutational Multivariate Analysis of Variance test. 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. 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 (triangles: cellulose group, circles: 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).

representative temporal trends (Fig. 3A, Fig. S6). With this coarse-grained data representation, we then obtained two p values by comparing the differential responses of the extracted features between the intervention and control group (“responsiveness”, p_r) as well as those between the four vendors in the intervention group (“baseline dependence”, p_b) using Permutational Multivariate Analysis of Variance (PERMANOVA) test. The NMF-PERMANOVA analysis confirmed that the dynamical responses of bacterial load (Fig. 3B), propionate, and butyrate (Fig. 3C) were nontrivial and baseline-dependent (both p_r and $p_b < 0.05$ 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 the mouse gut microbiota in response to dietary fiber intervention (Fig. 4A, see “Methods”). The gLV model assumes that degradation and subsequent utilization of dietary fibers boost bacterial growth rate (the amount of increment is parameterized by ϵ). One long-standing challenge in gLV inference is the overfitting of parameters when the time series is too short and the number of samples at all timepoints is far less than that of gLV parameters. To address the challenge, we estimated the uncertainty associated with model parameters by formulating the gLV-based inference in a Bayesian framework which outputs posterior distributions of estimated parameters

[40], rather than point estimates in penalized regressions [41]. In our gLV-based probabilistic framework, any bacteria taxa with a significantly positive distribution of “fiber response” ϵ is predicted as a “primary degrader” of inulin as gLV controls for inter-taxa interactions during regression. We identified five such primary degraders grouped at varying taxonomic levels, including *Muribaculaceae* (family), *Faecalibaculum* (genus), *Parasutterella* (genus), and *Bacteroides* (genus) and *Bacteroides acidifaciens* (species) (Fig. 4B, see “Methods”). The inference of primary degraders in the gLV-based framework was robust to the criteria used for clustering 16S rRNA gene amplicons, either at the lowest classified taxonomic level (Fig. 4B) or at the operational taxonomic unit (OTU, 97% sequence similarity) level (Fig. S7). In addition, we explored the use of linear mixed-effects (LMM) model to identify a significant association between bacteria taxa and diet (see “Methods”). We found that four of the five primary degraders inferred by gLV were also identified by LMM (Table S1).

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 S2). For example, members from *Bacteroides* and *Muribaculaceae* contain PULs with a *susC/susD* homologous gene pair that facilitates sensing and import of inulin [42, 43]. Putative inulin PULs were also detected in the metagenome-assembled genomes of *B. acidifaciens* and *Muribaculaceae* (Table S3). Furthermore, we analyzed the data from an independent

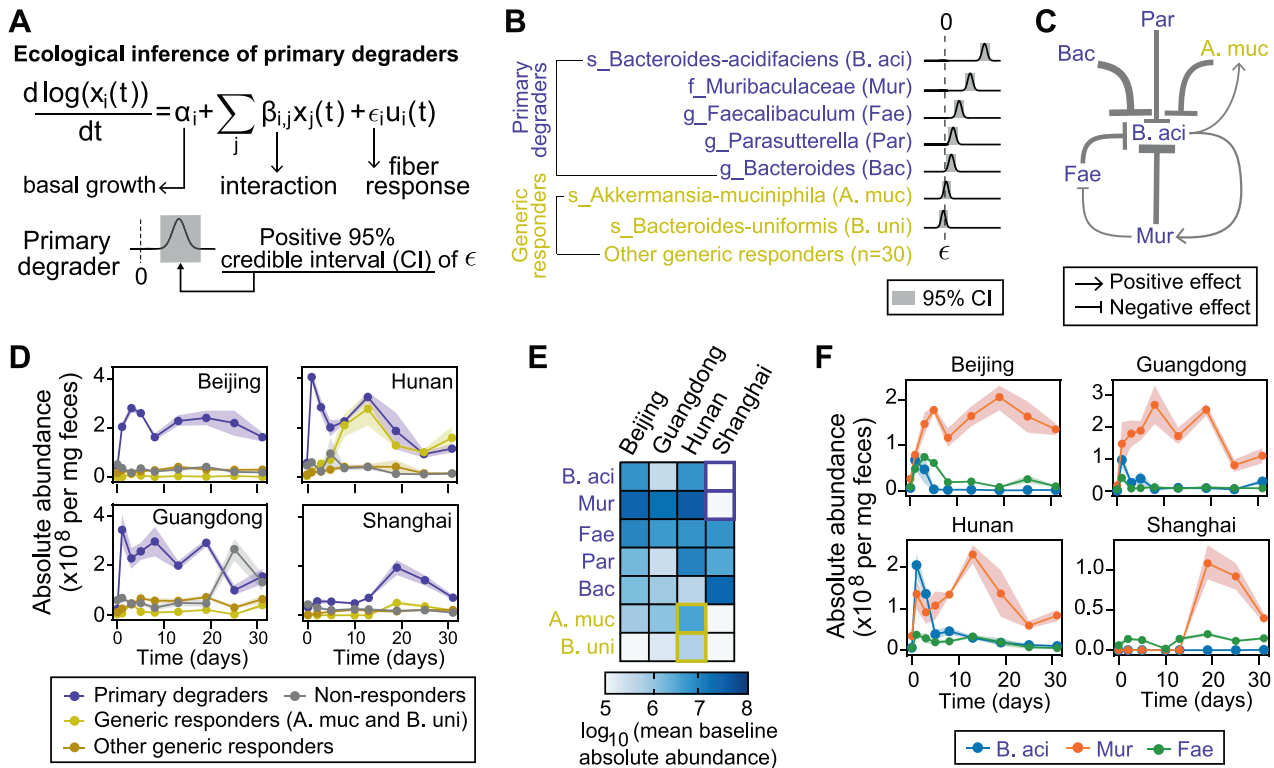


Fig. 4 Inulin responders shape the dynamical response of murine gut microbiota. **A** Generalized Lotka–Volterra (gLV) model combined with Bayesian statistics to infer inulin degraders and pairwise interactions. The gLV model summarizes the underlying ecology by three terms that additively determine bacterial growth rates: the basal growth rates (α), the influences from other bacteria (β), and the impacts of dietary fiber (ϵ). A primary degrader is determined when 95% credible interval of the posterior distribution of ϵ is positive. **B** Posterior distribution of ϵ 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 ϵ . **C** Core ecological interaction network composed of six bacterial taxa shown in panel B (see all significant interactions in Table S4). 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 (B). **F** Temporal changes in the absolute abundance of the top three inulin degraders. In **D** and **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.

study [44], 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 abundance was not available in this data set, we applied the Bayesian version of gLV-based inference to the relative abundance profiles and again identified *B. acidifaciens* as a primary degrader of inulin (Fig. S8C).

Alternatively, we used the NMF-PERMANOVA analysis (Fig. 3A) to identify all bacterial taxa that exhibited differential responses between inulin and cellulose groups, regardless of whether the responses are direct or indirect. We found a total of 37 bacterial taxa with significant dynamical responses (Table S1), including the five putative primary degraders as inferred by the gLV model above. Among the remaining 32 taxa, *Akkermansia muciniphila* and *Bacteroides uniformis* are the most abundant and their relative abundances significantly increased after day 5 (Figs. S9 and S10). Unlike gLV and LMM, NMF-PERMANOVA did not control for indirect effects on bacterial growth via ecological interactions. It is therefore likely that *A. muciniphila* and *B. uniformis* may indirectly benefit from the primary degraders (we call them “generic responders”). Indeed, inference of the gLV model suggested that the growth of *A. muciniphila* was facilitated by *B. acidifaciens* (Fig. 4C). This is also consistent with previous observations that *A.*

muciniphila cannot grow on inulin but can be significantly promoted by shorter chain fructo-oligosaccharides [45].

The notion of eco-group (or guild, i.e., a set of bacterial taxa that perform similar functions, is very useful to understand microbial ecology [26, 46]. 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 *Akkermansia muciniphila* and *Bacteroides 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 SCFA productions (Figs. S5 and S10).

Our gLV-based inference suggested strong competition among primary degraders, where *Muribaculaceae* inhibited the growth of

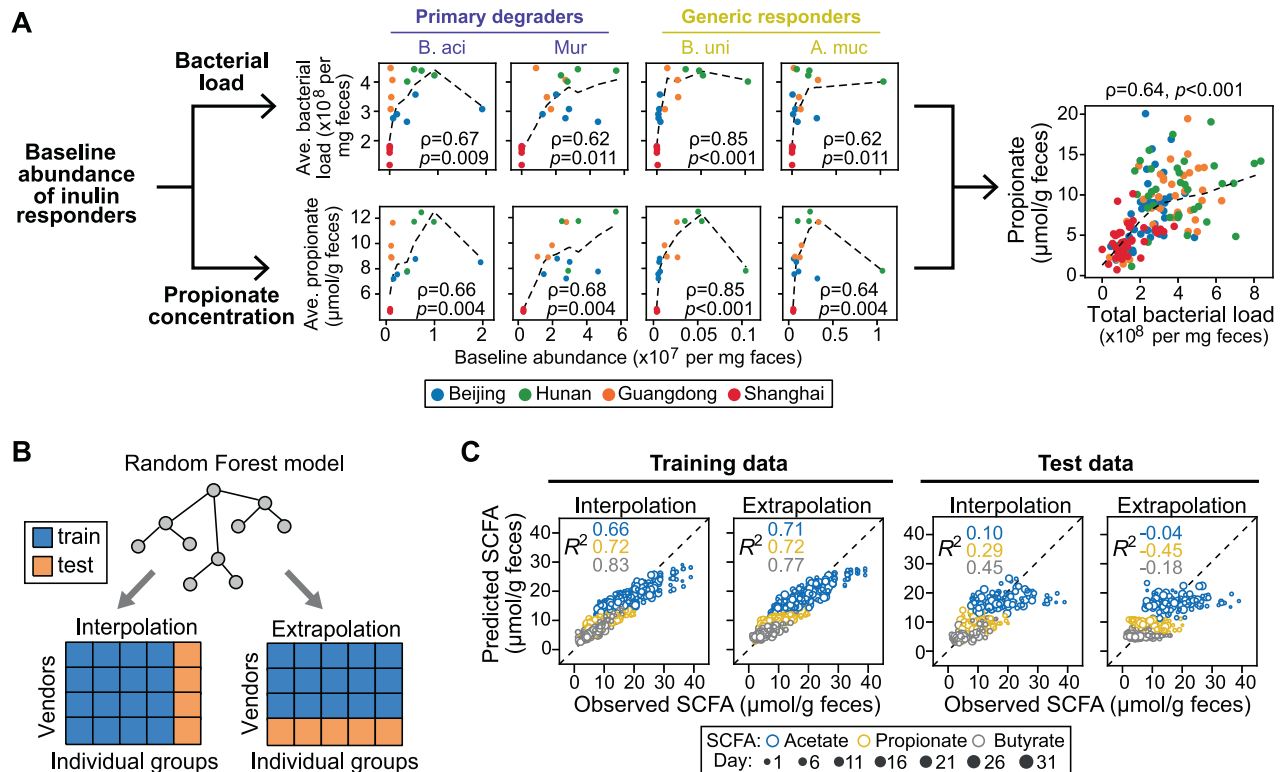


Fig. 5 Quantitative relationship between SCFAs and murine gut microbiota composition. **A** Correlation between bacterial load and propionate concentration (right big panel). We proposed that the correlation is mediated by some inulin responders which causally and simultaneously affect both observations. Eight small panels to the left: Spearman's correlations of baseline abundance of four inulin responders (*B. aci*: *s_Bacteroides-acidifaciens*, *Mur*: *f_Muribaculaceae*, *B. uni*: *s_Bacteroides-uniformis*, *A. muc*: *s_Akkermansia-muciniphila*) with the mean bacterial load (top row) or propionate concentration (bottom row) averaged across the intervention period. Dashed line: Lowess (Locally Weighted Scatterplot Smoothing) 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.

B. acidifaciens and *Facaelibaculum* (Fig. 4C). Indeed, *B. acidifaciens* and *Facaelibaculum* 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 (Figs. 4F and S10). Our results are consistent with previous studies by Patnode et al. [47] that identified competitive inhibition as the ecological mechanism for consistent dominance of *Bacteroides cellulosilyticus* over *Bacteroides vulgatus*, even though both species contain fiber-processing polysaccharide utilization loci (PULs). Taken together, we demonstrate that primary degraders and their 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 (Figs. 2A and 4D), due to the 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. muciphila*, and *B. uniformis* were positively correlated with the propionate concentration (Fig. 5A, left panel). Indeed, *Muribaculaceae*, *A. muciphila*, and *B.*

uniformis have been previously found to produce propionate in vitro and/or in vivo (Table S2). As a result and consistent with a previous study [48], 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 SCFA 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 the absolute abundance of bacterial taxa as predictors. All mice in our experiments were split into training data and test data using a 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 a 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 regression model fitted the training data reasonably well ($R^2 \geq 0.66$ regardless of SCFAs and data-split approaches), the predictions generalized poorly to the test data: R^2 of SCFAs ranged from 0.10 to 0.45 for "interpolation"

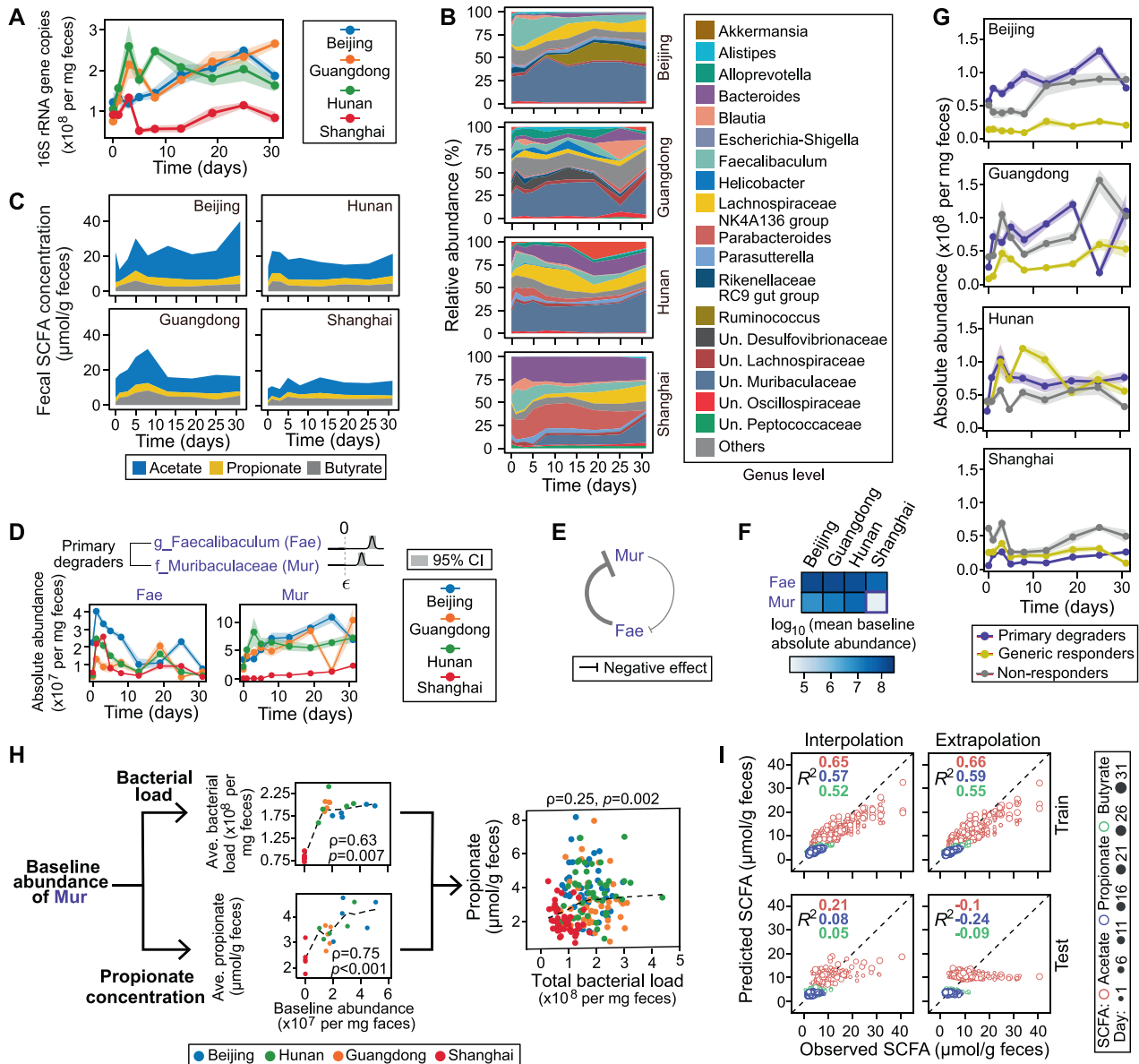


Fig. 6 Resistant starch-induced dynamical response in murine gut microbiota. Dynamical responses of bacterial load (A), gut microbiota composition (B), and SCFAs concentration (C) following the resistant starch intervention. D Dynamics of two putative resistant starch degraders. ϵ represents the growth impact of resistant starch and its posterior distributions are shown for each degrader. CI credible interval. E Ecological interactions between the two degraders (see all significant interactions in Table S4). 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. For the two small panels on the left, the time-averaged bacterial load or propionate concentration on the y-axis was calculated by dividing the area under the curve of the corresponding variables by the duration of observation. Dashed lines: Lowess (Locally Weighted Scatterplot Smoothing) regression. I Prediction of SCFAs concentration from gut microbiota composition using a Random Forest regression model. “Interpolation” and “extrapolation” are two strategies for splitting all data into the training and test sets (the same as Fig. 5B). Lines (A, D, G) or height of stacked bands represent mean values across mice from the same vendor ($n=5$ for all vendors). Shading areas (A, D, G) represent the standard error of the mean.

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). Given the current sample size, we found that Random Forest regression 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 human findings [49] that the substantial inter-individual

variation of gut microbiome could impede the predictive power of machine learning models.

Ecological response of murine gut microbiome to resistant starch

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

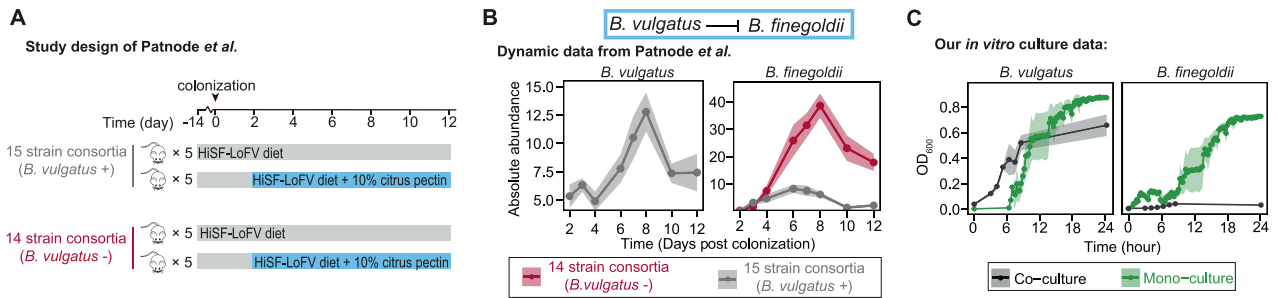


Fig. 7 Inference and validation of ecological interactions in synthetic human gut microbiome under citrus pectin intervention. **A** The study design of Patnode et al. [48]. In their study, synthetic communities of 14 human gut bacterial strains with and without *B. vulgatus* were continuously monitored in gnotobiotic mice for 12 days. Citrus pectin was supplemented after day 2 in the treatment group. **B** Time-series data of absolute abundance of *B. vulgatus* and *B. finegoldii* reported in the original study. Lines: mean; shading areas: standard error ($n = 5$ – 10 mice per group). **C** Growth of *B. finegoldii* and *B. vulgatus* in monoculture and co-culture experiments using media supplemented with citrus pectin (1% w/v). In the co-culture experiment, strains were inoculated at a 1:1 initial ratio. The abundance of individual strains in co-cultures was determined by qPCR.

starch stimulated milder changes in the bacterial load (Fig. 6A), gut microbiota composition (Fig. 6B), and SCFAs production (Fig. 6C). Under resistant starch intervention, different analyses (gLV, LMM, and NMF-PERMANOVA) all identified *Faecalibaculum* and *Muribaculaceae* as “primary degraders” (Fig. 6D, see Table S2 for evidence from literature) and NMF-PERMANOVA further detected 25 additional bacterial taxa as “generic responders” (Table S1). 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) exhibited 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 a 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 S1).

Finally, we found that the baseline abundance of *Muribaculaceae* was separately correlated with bacterial absolute abundance and propionate 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 composition 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.

Inference and validation of interspecies interactions in human gut microbiome

To further evaluate the generalizability of the gLV-based approach, we analyzed the ecological response of synthetic and natural human gut microbiome to dietary fiber from previously published time-series data, and validated the inferred interspecies interactions through in vitro experiments.

To start with, we applied the gLV framework to the longitudinal data from the study by Patnode et al. [47] and validated a

previously unidentified interspecies interaction. In their study, the authors profiled the ecological dynamics of synthetic communities of human gut bacteria strains in gnotobiotic mice under citrus pectin intervention (Fig. 7A). Consistent with the findings in their study, the gLV model successfully identified the inhibition of *B. vulgatus* by *B. cellulosilyticus* (Table S5). In addition, based on the longitudinal profiling of 14-strain (*B. vulgatus*–) and 15-strain (*B. vulgatus*+) baseline communities (Fig. 7B), we inferred that *B. vulgatus* is a strong competitor to *Bacteroides finegoldii*, which was not noted in the original study. We performed pairwise co-culture of these two species (see “Methods”), and found that the growth of *B. finegoldii* was strongly inhibited in the presence of *B. vulgatus* (Fig. 7C).

We then used the gLV framework to infer ecological drivers and interactions in natural human gut microbiome, based on longitudinal profiling of a human cohort under dietary fiber intervention from a previously published study by Baxter et al. [12] (Fig. 8A). We identified *Bifidobacterium adolescentis* as a strong inulin responder, which is consistent with the literature and validated by our experiment in vitro (Fig. 8B). Furthermore, we predicted a competitive interaction between *B. adolescentis* and *Bacteroides stercoris* (Table S6). In pairwise co-culture experiments of these two species, we observed that the growth of *Bacteroides stercoris* was inhibited in the presence of *B. adolescentis* (Fig. 8C).

Finally, we added the strong competitor *Bifidobacterium adolescentis* to different baseline communities lacking it and evaluated the outcome (Fig. 8D). In the baseline communities, which were derived from fecal samples of three healthy volunteers, *Bifidobacterium adolescentis* was absent and *Bacteroides stercoris* was present. We used the MiPro model for in vitro culture, which has been shown to maintain the compositional profiles of individual gut microbiomes (see “Methods”) [50]. After 24 h of in vitro culture, we sequenced the communities and evaluated the effect of adding *B. adolescentis* (Fig. 8E). With a significant increase of *B. adolescentis* in the community, we found that the growth of *Bacteroides stercoris* was inhibited in two of the three baselines.

In summary, we have shown that the gLV model can be used as a useful framework to infer ecological drivers and interactions in complex gut microbial communities, leading to testable hypotheses for experimental validation [24].

DISCUSSION

Our study characterizes the ecosystem response of murine gut microbiota to fibers, and emphasizes that ecological interactions play a key role in the personalized impact of dietary changes. gLV-based ecological inference from gut microbiome time-series data has yielded mechanistic insights into the stability of probiotic

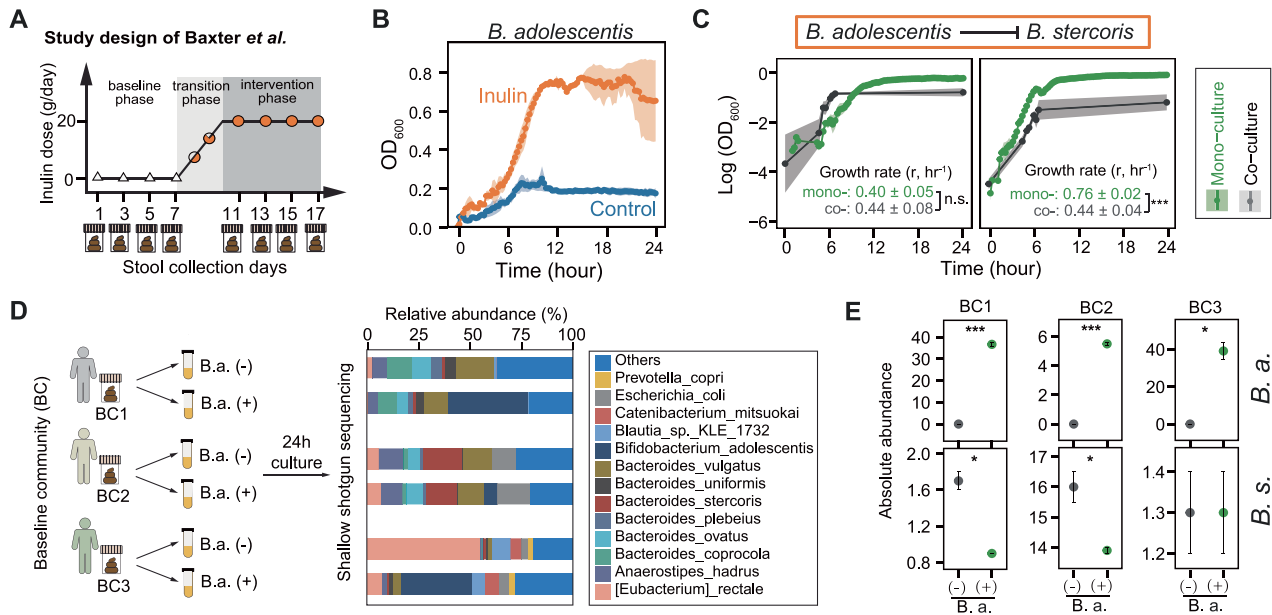


Fig. 8 Inference and validation of ecological driver and interactions in human gut microbiome under inulin intervention. **A** The study design of Baxter et al. [12]. In their study, the gut microbiota of 49 healthy volunteers was tracked before and during inulin intervention for more than two weeks. **B** Growth of *B. adolescentis* with or without inulin ($n = 3$). **C** Growth of *B. stercoris* and *B. adolescentis* in monoculture and co-culture experiments. In the co-culture experiment, strains were inoculated at a 1:1 initial ratio in media supplemented with inulin. The abundance of individual strains in co-cultures was determined by qPCR. **D** *B. adolescentis* was added to three baseline microbial communities derived from fecal samples of healthy volunteers. After in vitro culture of 24 h, the taxonomic profile of the microbial communities was determined by metagenomic sequencing and shown in horizontal stacked bars. *B.a* (-): *B. adolescentis* was not added in the baseline communities; *B.a* (+): *B. adolescentis* was added in the baseline communities. **E** Absolute abundance (unit: $\text{OD}_{600} \times 10^{-2}$) of *B. adolescentis* (*B. a.*) and *B. stercoris* (*B.s.*) after in vitro culture of 24 h. Dots: mean; bars: standard error. p values were computed using t-test ($n = 3$): *** $p < 0.001$; * $p < 0.05$.

community under dietary perturbation [51], colonization resistance of pathogenic *Clostridioides difficile* [52], and community assembly dynamics within preterm infant gut [24]. By integrating gLV model with Bayesian regression, we inferred a competitive network of fiber degraders as key bacteria that mediate the response of murine gut microbiota to inulin and resistant starch intervention. Besides evidences supporting the fiber-degrading function of putative degraders, our study confirms findings in the literature and advances the understanding of the effects of dietary fibers on the gut microbiome at the system level. 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 in a restricted number of taxa. Second, the absolute abundance of many fiber-degrading bacteria such as taxa related to the genus *Bifidobacterium*, failed to expand in the mouse gut on both fibers (Fig. S15), indicating that fiber-induced bacterial enrichment cannot be simply predicted from their in vitro growth. Third, we reasoned that the personalized fiber-induced response of gut microbiota was largely determined by the baseline abundance of fiber degraders and the ecological interactions among these degraders. In our study, the inter-individual variability of fiber-induced shift in propionate production was associated with the baseline abundance of a key fiber degrader *Muribaculaceae*. This is consistent with previous findings that the abundance of a key starch-degrader prior to resistant starch intervention was indicative of whether an individual would have a higher fecal butyrate response [16]. The ecological interactions among fiber degraders, on the other hand, drive the cascading alterations during the intervention. Our results revealed that fiber-induced dynamics of murine gut microbiota were largely driven by competitions and *Muribaculaceae* out-competed other degraders in consuming both inulin and resistant starch. Since the family *Muribaculaceae* was specific to the mouse

gut [53], it might have been adapted to the murine gut with higher fitness.

We have shown that the inference of gLV model can be experimentally validated, however, the generalizability of inferred interactions should be taken with caution in contexts that differ from the data used for inference, such as in different communities, and diet regimes, etc. For example, while we observed a clear pairwise interaction in the co-culture experiment, we found that *Bacteroides stercoris* was not inhibited in one of the three baseline communities with the addition of *B. adolescentis* (Fig. 8E). One possible explanation is the indirect effect mediated by other species [54]; another explanation is higher-order interactions [55], which are not captured by our current gLV framework. Regarding to diet regimes, we found that the competitive interactions that we validated by in vitro co-culture experiments still hold in the absence of dietary fiber supplementation (Fig. S16). Conceptually, whether the interspecies interactions are specific to diets would depend on the underlying biological mechanisms. For example, we may expect competitive interactions mediated by T6SS or secreted secondary metabolites to be insensitive to the type of nutrients provided; however, competitive interactions mediated by nutrient utilization are more likely to differ across dietary regimes. Despite these caveats, we have demonstrated the utility of gLV inference to generate testable hypotheses of key species and interspecies interactions [24].

Understanding the microbial responses and its association with the baseline microbiota composition rightly is a critical step in individualized dietary fiber intervention. To date, most of the current studies are based on cross-sectional study design and described the pre-to-post changes in abundance/concentration as microbial responses to dietary intervention [13, 15]. However, it should be noted that gut microbiome is a highly dynamic ecosystem, and its response to dietary fibers could have temporal

characteristics [38]. Consequently, the significance of microbial responses 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. S17). Furthermore, due to the lack of control group data to assess the intervention effects, pre-to-post changes that are supposed to capture fiber-induced effects may be entirely attributable to random temporal variations within each individual [56]. We speculate that these two caveats are the main cause of “reproducibility crisis” [57] among microbiome researches. In contrast, our analysis avoids the above two caveats by incorporating longitudinal data and a control group. Additionally, the use of dimensionality reduction in our approach further benefits data visualization of inter-vendor variations in gut microbiota composition (Fig. 3).

Diet-induced changes in SCFAs are often transient and vanish shortly after cessation of dietary intervention [9, 58–60]. Our study is consistent with this result, by showing that SCFA concentrations cannot be maintained at their 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 [61] 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 [62]. While many studies used fecal SCFAs concentrations as a 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. We note that dietary fibers (e.g., inulin) may have direct effects on host properties, including absorption of SCFA [63], regulation of host mucosal signaling [64], etc. Due to the difficulty of measuring fluxes *in vivo*, mathematical models that take both processes into account show great promise in the estimation of their flux rates from SCFA concentrations [65].

Characterizing the dynamics of gut microbiota and their inter-individual variability with multi-omics data is an important priority for microbiome research to further understand diet-induced responses [66]. Such studies hold great promise to improve human health and treat gut microbiome-associated diseases 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 the dietary response of a gut bacterial taxa depends on the ecological context [67]. By inventing a new application of gLV with uncertainty assessment to infer primary fiber degraders and the associated interaction network, we provide a generalizable computational approach to study the ecological dynamics of the gut microbiome under dietary interventions. Systems biology modeling, both top-down and bottom-up (e.g., genome-scale metabolic models) approaches, is much needed to understand the functional outputs of microbiome [68–70]. 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 [71].

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 a 12-h light/dark cycle and allowed *ad libitum* access to food and water throughout the experiment. After acclimatizing to the cellulose-based 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 ($n = 5$). During the intervention period, mice in the cellulose group were maintained on the cellulose-based diet, while mice in the fiber intervention groups switched to a diet supplemented with inulin (inulin group) or resistant starch (RS group). Composition of all diets including the source of dietary fibers cellulose, resistant starch from maize (HI-MAIZE 260, Ingredion Inc.), and inulin (Orafti HP, BENE0-Orafti) are provided in Table S7. Fecal pellets from each mouse were freshly collected over multiple time points: day 0 (before diet change), day 1, 3, 5, 8, 13, 19, 25, and 31 (Fig. 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 (SIAT-IACUC-190606-HCS-LHB-A0754).

Collection of human fecal samples

Fecal samples were collected from healthy volunteers. Fresh stool samples were collected in sterile PBS pre-reduced with 0.1% (w/v) L-cysteine hydrochloride. The samples were immediately weighed and transferred into an anaerobic workstation (5% H_2 , 5% CO_2 , and 90% N_2 at 37°C). Then, samples were homogenized with a vortex mixer and filtered using sterile gauzes. Aliquots were placed in sterile cryogenic vials and frozen at -80°C until use. The protocol was approved by the Institutional Review Board/Ethics committees at the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences (#YSB-2022-Y02043).

In vitro culture of human gut bacteria strains

Bacteria were grown overnight in BHI media to the early stationary phase. Bacteria cells were centrifuged and washed with fresh MiPro media before inoculation. Pectin from citrus peel was purchased from Sigma-Aldrich (St. Louis, MO, USA). Citrus pectin or inulin was added to the MiPro media [50] at a final concentration of 1% (g/ml). Bacteria strains were cultured in an anaerobic chamber (Coy, specify model; 85% N_2 , 10% H_2 , and 5% CO_2) at 37°C . In the monoculture experiment, the initial OD_{600} of bacterial cells is 0.01; in pairwise co-culture experiments, the initial OD_{600} of both strains is 0.01 (i.e. total OD_{600} is 0.02). Bacterial growth in mono-/co-culture experiments was measured in 96-well plates using a GenTech Epoch2 plate reader. Strains of *B. adolescentis*, *B. stercoris*, *B. vulgatus*, and *B. fingoldii* were isolated from healthy human stool samples. The whole-genome sequence was used for taxonomic classification based on Genome Taxonomy Database Toolkit (GTDB-tk) [72].

In vitro culture of human gut microbial communities

Three samples were chosen as baseline communities as they met the following criteria: (1) absence of *Bifidobacterium*; (2) presence of *Bacteroides stercoris*. The communities were cultured in 96-well plates using MiPro media supplemented with inulin to examine the ecological response. Stock solutions of inulin were sterilized by membrane filtration (Millex-GP filters with 0.22- μm pores, Millipore) and added aseptically to the MiPro media at a final concentration of 1% (g/ml). The baseline communities (20 μl) were thawed and inoculated into the wells containing MiPro medium (980 μl). Overnight culture of *B. adolescentis* was centrifuged, resuspended, and added to the baseline communities ($\sim 4 \times 10^7$ CFU of *B. adolescentis* per well). The communities were cultured in an anaerobic workstation for 24 h before sequencing.

Quantification of fecal SCFA concentration by GC-MS

The SCFAs of mice fecal samples were analyzed by GC-MS [73, 74]. For the sample extraction, 0.05 g of frozen feces were mixed with 300 μl of pure

water containing caproic acid-6,6-d₃ (CDN Isotopes, Quebec, Canada) as internal standard (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 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 was accurately transferred into a glass insert in a GC vial and capped tightly after adding 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, iso-butyric, iso-valeric, valeric, 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 integrated areas for all SCFAs were normalized with the internal standard and quantified with the standard curve, as previously described [74].

DNA extraction of samples from in vitro culture experiments

For co-culture samples, DNA extraction was performed in the 96-well format using the TIAMap Bacteria DNA Kit (TIANGEN Biotech, Beijing) following the corresponding lysis protocol for Gram-positive bacterial cells using lysozyme and proteinase K. The genomic DNA of in vitro microbial community culture samples were isolated using a DNeasy UltraClean microbial kit (Qiagen, Hilden, Germany) and quantitated using the Qubit fluorometer (ThermoFisher Scientific).

DNA extraction of fecal samples 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 the bacterial load, total bacteria density was determined using qPCR as previously described [75]. A 466-bp fragment of the bacterial 16S rRNA gene was amplified using the forward primer 5'-TCCTACGGGAGGCAGCAGT-3' and the reverse primer 5'-GGACTACCAGGGTATCTAATCCTGTT-3'. The absolute abundance of a bacterial taxon was estimated by the multiplication of its relative abundance and the total bacterial load.

Quantitative PCR of bacterial abundance in co-culture samples

Co-culture samples were processed for gDNA extraction and qPCR with species-specific primers (Table S8). To quantify the abundance of the two-bacterial species in co-culture, we performed quantitative PCR using a CFX96 Real-Time System (Bio-Rad Laboratories) as previously described [76]. Briefly, samples were analyzed in a 20-µl reaction mix consisting of 10 µl 1×SYBR Green Master Mix buffer (Takara, Dalian, China), 0.1 µM of species-specific primer, and 3 µl of template DNA. Overnight single-strain cultures were diluted with a twofold gradient to construct the standard curve. The abundance of bacterial strains in the co-culture was estimated based on the standard curves.

16S rRNA gene 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. The extracted DNA sample was

purified using silica-based columns. Metagenomic sequence 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 gene sequencing reads were analyzed by QIIME2 (version 2020.2) [79]. Demultiplexed paired-end reads were trimmed to remove primers and low-quality bases with the 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 [80]. The generated feature table was filtered to remove ASVs present in only a single sample and the 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, 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 a great majority of the bacteria species diversity and richness that could be sampled was captured by our sequencing depth (Fig. S18), which indicated sufficient sequencing depth for the 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) [33]. 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 based on the classify-sklearn naïve Bayes taxonomy classifier was used with the SILVA (v.138) as the reference database. Unless specified (Figs. 2C, 6B and Figs. S1A, S7), ASVs were grouped to the lowest classified taxonomy level (i.e., grouping 16S rRNA gene sequencing at a specified taxonomic level, excluding those classified at lower levels) for all data modeling and analysis. The specified taxonomic levels are labeled with the 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. Alternatively, ASV sequences were grouped into operational taxonomic units (OTUs) at 97% similarity (Fig. S7).

For metagenome analysis, raw sequencing reads were subjected to quality filtering and barcode trimming using KneadData (v0.5.4) by employing trimmomatic settings of a 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 the 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 catalog 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: $RPM = 1 M \times (\text{mapped reads/gene length}) / (\text{sum of mapped reads/gene length})$ [87]. For all the predicted genes, CAZymes were annotated using hmsearch 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 (<2000 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 [72], and subjected

where $x_{i,k} = x_i(t_k)$ and $u_k = u(t_k)$. The log-derivatives of x_i on the left-hand side of Eq. (3) were estimated from a cubic spline interpolation. Using a simplified notation for Eq. (3), i.e., $\mathbf{Y}_i = \mathbf{X}_i \mathbf{C}_i$, we can incorporate data from all mice into a single regression model

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_p \end{bmatrix} = \begin{bmatrix} X_r & 0 & 0 \\ \cdots & \cdots & \cdots \\ 0 & 0 & X_p \end{bmatrix} \begin{bmatrix} C_1 \\ \vdots \\ C_p \end{bmatrix} \quad (4)$$

The linear regression as described in Eq. (4) (for brevity $\mathbf{Y} = \mathbf{X}\mathbf{C}$) can be further transformed into a Bayesian regression $\mathbf{Y} = \mathcal{N}(\mathbf{X}\mathbf{C}, \sigma)$ where \mathcal{N} and σ 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. For all gLV inferences, the time-series data from all mice in the treatment group and those in the control group (as long as the control group data is available) were combined in one parameter fitting based on the hypothesis that gLV parameters do not change their values between diets [95]. We used uninformative Normal priors $\mathcal{N}(0, 1)$ for all gLV parameters and Stan program [96] 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. Our Bayesian approach is conceptually similar to the Bayesian adaptive lasso algorithm in MDSINE [51]; the key difference is that MDSINE uses a hierarchical probability model and regularization to sample the variance parameters in the Normal priors.

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, the pipeline was applied to the training data only and five hyperparameters were tuned using fivefold cross-validation within the training set and R^2 as the scoring metric (GridSearchCV function of the scikit-learn library in Python): 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 minimum 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 2000.

DATA AVAILABILITY

All data (including metadata, sequencing data, and metabolomics data) have been deposited in the NCBI SRA under accession number PRJNA754019.

CODE AVAILABILITY

The customized scripts used in this study are available at: <https://github.com/liaochen1988/DFdynamics>.

REFERENCES

- Tolhurst G, Heffron H, Lam YS, Parker HE, Habib AM, Diakogiannaki E, et al. Short-chain fatty acids stimulate glucagon-like peptide-1 secretion via the G-protein-coupled receptor FFAR2. *Diabetes*. 2012;61:364–71.
- Vinolo MAR, Rodrigues HG, Nachbar RT, Curi R. Regulation of inflammation by short chain fatty acids. *Nutrients*. 2011;3:858–76.
- Litvak Y, Byndloss MX, Bäumlér AJ. Colonocyte metabolism shapes the gut microbiota. *Science*. 2018;362:t9076.
- Sanna S, van Zuydam NR, Mahajan A, Kurilshikov A, Vich Vila A, Vösa U, et al. Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. *Nat Genet*. 2019;51:600–5.
- Parada Venegas D, De la Fuente MK, Landskron G, González MJ, Quera R, Dijkstra G, et al. Short chain fatty acids (SCFAs)-mediated gut epithelial and immune regulation and its relevance for inflammatory bowel diseases. *Front Immunol*. 2019;10:277.
- Zhao L, Zhang F, Ding X, Wu G, Lam YY, Wang X, et al. Gut bacteria selectively promoted by dietary fibers alleviate type 2 diabetes. *Science*. 2018;359:1151–6.
- Sitkin S, Vakhitov T, Pokrotnieks J. How to increase the butyrate-producing capacity of the gut microbiome: do IBD patients really need butyrate replacement and butyrogenic therapy? *J Crohn's Colitis*. 2018;12:881–2.
- Lordan C, Thapa D, Ross RP, Cotter PD. Potential for enriching next-generation health-promoting gut bacteria through prebiotics and other dietary components. *Gut Microbes*. 2019;11:1–20.
- David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature*. 2014;505:559–63.
- Singh V, Yeoh BS, Walker RE, Xiao X, Saha P, Golonka RM, et al. Microbiota fermentation-NLRP3 axis shapes the impact of dietary fibres on intestinal inflammation. *Gut*. 2019;68:1801–12.
- Healey G, Murphy R, Butts C, Brough L, Whelan K, Coad J. 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. *Brit J Nutr*. 2018;119:176–89.
- Baxter NT, Schmidt AW, Venkataraman A, Kim KS, Waldron C, Schmidt TM. Dynamics of human gut microbiota and short-chain fatty acids in response to dietary interventions with three fermentable fibers. *mBio*. 2019;10:e02566–18.
- Deehan EC, Yang C, Perez-Muñoz ME, Nguyen NK, Cheng CC, Triador L, et al. Precision microbiome modulation with discrete dietary fiber structures directs short-chain fatty acid production. *Cell Host Microbe*. 2020;27:389–404.
- Venkataraman A, Sieber JR, Schmidt AW, Waldron C, Theis KR, Schmidt TM. Variable responses of human microbiomes to dietary supplementation with resistant starch. *Microbiome*. 2016;4:33.
- Nguyen NK, Deehan EC, Zhang Z, Jin M, Baskota N, Perez-Muñoz ME, 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;8:118.
- Ze X, Duncan SH, Louis P, Flint HJ. *Ruminococcus bromii* is a keystone species for the degradation of resistant starch in the human colon. *ISME J*. 2012;6:1535–43.
- Lahti L, Salojärvi J, Salonen A, Scheffer M, de Vos WM. Tipping elements in the human intestinal ecosystem. *Nat Commun*. 2014;5:4344.
- Rodríguez J, Hiel S, Neyrinck AM, Le Roy T, Pötgens SA, Leyrolle Q, et al. Discovery of the gut microbial signature driving the efficacy of prebiotic intervention in obese patients. *Gut*. 2020;69:1975–87.
- Kovatcheva-Datchary P, Nilsson A, Akrami R, Lee YS, De Vadder F, Arora T, et al. Dietary fiber-induced improvement in glucose metabolism is associated with increased abundance of *Prevotella*. *Cell Metab*. 2015;22:971–82.
- Coyte KZ, Schluter J, Foster KR. The ecology of the microbiome: networks, competition, and stability. *Science*. 2015;350:663–6.
- Davis LMG, Martínez I, Walter J, Goin C, Hutkins RW. Barcoded pyrosequencing reveals that consumption of galactooligosaccharides resulted in a highly specific bifidogenic response in humans. *PLoS ONE*. 2011;6:e25200.
- Solden LM, Naas AE, Roux S, Daly RA, Collins WB, Nicora CD, et al. Interspecies cross-feeding orchestrates carbon degradation in the rumen ecosystem. *Nat Microbiol*. 2018;3:1274–84.
- Rakoff-Nahoum S, Coyne MJ, Comstock LE. An ecological network of polysaccharide utilization among human intestinal symbionts. *Curr Biol*. 2014;24:40–9.
- Rao C, Coyte KZ, Bainter W, Geha RS, Martin CR, Rakoff-Nahoum S. Multi-kingdom ecological drivers of microbiota assembly in preterm infants. *Nature*. 2021;591:633–8.
- Koskella B, Hall LJ, Metcalf C. The microbiome beyond the horizon of ecological and evolutionary theory. *Nat Ecol Evol*. 2017;1:1606–15.
- Goldford JE, Lu N, Bajic D, Estrela S, Tikhonov M, Sanchez-Gorostiaga A, et al. Emergent simplicity in microbial community assembly. *Science*. 2018;361:469–74.
- Ortiz A, Vega NM, Ratzke C, Gore J. Interspecies bacterial competition regulates community assembly in the *C. elegans* intestine. *ISME J*. 2021;15:2131–45.
- Liu Z, de Vries B, Gerritsen J, Smidt H, Zoetendal EG. Microbiome-based stratification to guide dietary interventions to improve human health. *Nutr Res*. 2020;82:1–10.
- Ahmed W, Rashid S. Functional and therapeutic potential of inulin: a comprehensive review. *Crit Rev Food Sci Nutr*. 2019;59:1–13.

30. Cerqueira FM, Potenhauer AL, Pollet RM, Brown HA, Koropatkin NM. Starch digestion by gut bacteria: crowdsourcing for carbs. *Trends Microbiol.* 2019;28:95–108.
31. Parker KD, Albeke SE, Gigley JP, Goldstein AM, Ward NL. Microbiome composition in both wild-type and disease model mice is heavily influenced by mouse facility. *Front Microbiol.* 2018;9:1598.
32. Ericsson AC, Davis JW, Spollen W, Bivens N, Givan S, Hagan CE, et al. Effects of vendor and genetic background on the composition of the fecal microbiota of inbred mice. *PLoS ONE.* 2015;10:e116704.
33. Martino C, Morton JT, Marotz CA, Thompson LR, Tripathi A, Knight R, et al. A novel sparse compositional technique reveals microbial perturbations. *mSystems.* 2019;4:e00016–19.
34. Lagkouravdos I, Lesker TR, Hitch TCA, Gálvez EJC, Smit N, Neuhaus K, et al. Sequence and cultivation study of Muribaculaceae reveals novel species, host preference, and functional potential of this yet undescribed family. *Microbiome.* 2019;7:28.
35. Pereira FC, Wasmund K, Cobankovic I, Jehmlich N, Herbold CW, Lee KS, et al. Rational design of a microbial consortium of mucosal sugar utilizers reduces *Clostridiodes difficile* colonization. *Nat Commun.* 2020;11:5104.
36. Almeida A, Mitchell AL, Boland M, Forster SC, Gloor GB, Tarkowska A, et al. A new genomic blueprint of the human gut microbiota. *Nature.* 2019;568:499–504.
37. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, et al. Linking long-term dietary patterns with gut microbial enterotypes. *Science.* 2011;334:105–8.
38. Creswell R, Tan J, Leff JW, Brooks B, Mahowald MA, Thieroff-Ekerdt R, et al. High-resolution temporal profiling of the human gut microbiome reveals consistent and cascading alterations in response to dietary glycans. *Genome Med.* 2020;12:59.
39. Mackevicius EL, Bahle AH, Williams AH, Gu S, Denisenko NI, Goldman MS, et al. Unsupervised discovery of temporal sequences in high-dimensional datasets, with applications to neuroscience. *Elife.* 2019;8:e38471.
40. Morjaria S, Schluter J, Taylor BP, Littmann ER, Carter RA, Fontana E, et al. Antibiotic-induced shifts in fecal microbiota density and composition during hematopoietic stem cell transplantation. *Infect Immun.* 2019;87:e00206.
41. Stein RR, Bucci V, Toussaint NC, Buffie CG, Ratsch G, Pamer EG, et al. Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. *PLoS Comput Biol.* 2013;9:e1003388.
42. Rakoff-Nahoum S, Foster KR, Comstock LE. The evolution of cooperation within the gut microbiota. *Nature.* 2016;533:255–9.
43. Koropatkin NM, Cameron EA, Martens EC. How glycan metabolism shapes the human gut microbiota. *Nat Rev Microbiol.* 2012;10:323–35.
44. Chijiwa R, Hosokawa M, Kogawa M, Nishikawa Y, Ide K, Sakanashi C, et al. Single-cell genomics of uncultured bacteria reveals dietary fiber responders in the mouse gut microbiota. *Microbiome.* 2020;8:5–14.
45. Zhou K. Strategies to promote abundance of *Akkermansia muciniphila*, an emerging probiotics in the gut, evidence from dietary intervention studies. *J Funct Foods.* 2017;33:194–201.
46. Wu G, Zhao N, Zhang C, Lam YY, Zhao L. Guild-based analysis for understanding gut microbiome in human health and diseases. *Genome Med.* 2021;13:22.
47. Patnode ML, Beller ZW, Han ND, Cheng J, Peters SL, Terrapon N, et al. Interspecies competition impacts targeted manipulation of human gut bacteria by fiber-derived glycans. *Cell.* 2019;179:59–73.
48. Salonen A, Lahti L, Salojärvi J, Holtrop G, Korpela K, Duncan SH, et al. Impact of diet and individual variation on intestinal microbiota composition and fermentation products in obese men. *ISME J.* 2014;8:2218–30.
49. Sze MA, Topçuoğlu BD, Lesniak NA, Ruffin MT, Schloss PD. Fecal short-chain fatty acids are not predictive of colonic tumor status and cannot be predicted based on bacterial community structure. *mBio.* 2019;10:e1419–54.
50. Li L, Abou-Samra E, Ning Z, Zhang X, Mayne J, Wang J, et al. An in vitro model maintaining taxon-specific functional activities of the gut microbiome. *Nat Commun.* 2019;10:4146.
51. Bucci V, Tzen B, Li N, Simmons M, Tanoue T, Bogart E, et al. MDSINE: Microbial Dynamical Systems Inference Engine for microbiome time-series analyses. *Genome Biol.* 2016;17:121.
52. Buffie CG, Bucci V, Stein RR, McKenney PT, Ling L, Goubourne A, et al. Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. *Nature.* 2015;517:205–8.
53. Lagkouravdos I, Pukall R, Abt B, Foesele BU, Meier-Kolthoff JP, Kumar N, et al. The Mouse Intestinal Bacterial Collection (miBC) provides host-specific insight into cultured diversity and functional potential of the gut microbiota. *Nat Microbiol.* 2016;1:16131.
54. Xiao Y, Angulo MT, Lao S, Weiss ST, Liu Y. An ecological framework to understand the efficacy of fecal microbiota transplantation. *Nat Commun.* 2020;11:3329.
55. Worthen WB, Moore JL. Higher-order interactions and indirect effects: a resolution using laboratory *Drosophila* communities. *Am Nat.* 1991;138:1092–104.
56. Atkinson G, Batterham AM. True and false interindividual differences in the physiological response to an intervention. *Exp Physiol.* 2015;100:577–88.
57. Schloss PD. Identifying and overcoming threats to reproducibility, replicability, robustness, and generalizability in microbiome research. *mBio.* 2018;9:e00525.
58. Baxter NT, Lesniak NA, Sinani H, Schloss PD, Koropatkin NM. The glucoamylase inhibitor acarbose has a diet-dependent and reversible effect on the murine gut microbiome. *mSphere.* 2019;4:e00528.
59. Walker AW, Ince J, Duncan SH, Webster LM, Holtrop G, Ze X, et al. Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *ISME J.* 2011;5:220–30.
60. Hiel S, Bindels LB, Pachikian BD, Kalala G, Broers V, Zamariola G, 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;109:1683–95.
61. Nordgaard I, Hove H, Clausen MR, Mortensen PB. Colonic production of butyrate in patients with previous colonic cancer during long-term treatment with dietary fibre (*Plantago ovata* seeds). *Scand J Gastroenterol.* 1996;31:1011–20.
62. Sakata T. Pitfalls in short-chain fatty acid research: a methodological review. *Anim Sci J.* 2019;90:3–13.
63. McNeil NI, Cummings JH, James WP. Short chain fatty acid absorption by the human large intestine. *Gut.* 1978;19:819–22.
64. Wu RY, Määttänen P, Napper S, Scruten E, Li B, Koike Y, et al. Non-digestible oligosaccharides directly regulate host kinome to modulate host inflammatory responses without alterations in the gut microbiota. *Microbiome.* 2017;5:135.
65. Gurry T, Nguyen L, Yu X, Alm EJ. Functional heterogeneity in the fermentation capabilities of the healthy human gut microbiota. *PLoS ONE.* 2021;16:e254004.
66. Johnson AJ, Zheng JJ, Kang JW, Saboe A, Knights D, Zivkovic AM. A guide to diet-microbiome study design. *Front Nutr.* 2020;7:79.
67. Shepherd ES, DeLoache WC, Pruss KM, Whitaker WR, Sonnenburg JL. An exclusive metabolic niche enables strain engraftment in the gut microbiota. *Nature.* 2018;557:434–8.
68. Kumar M, Ji B, Zengler K, Nielsen J. Modelling approaches for studying the microbiome. *Nat Microbiol.* 2019;4:1253–67.
69. Gowda K, Ping D, Mani M, Kuehn S. Genomic structure predicts metabolite dynamics in microbial communities. *Cell.* 2022;185:530–46.
70. Qian Y, Lan F, Venturelli OS. Towards a deeper understanding of microbial communities: integrating experimental data with dynamic models. *Curr Opin Microbiol.* 2021;62:84–92.
71. Kolodziejczyk AA, Zheng D, Elinav E. Diet-microbiota interactions and personalized nutrition. *Nat Rev Microbiol.* 2019;17:742–53.
72. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics.* 2019;36:1925–7.
73. Zhang S, Wang H, Zhu M. A sensitive GC/MS detection method for analyzing microbial metabolites short chain fatty acids in fecal and serum samples. *Talanta.* 2019;196:249–54.
74. Cai J, Zhang J, Tian Y, Zhang L, Hatzakis E, Krausz KW, et al. Orthogonal comparison of GC-MS and ¹H NMR spectroscopy for short chain fatty acid quantification. *Anal Chem.* 2017;89:7900–6.
75. Jian C, Luukkonen P, Yki-Järvinen H, Salonen A, Korpela K. Quantitative PCR provides a simple and accessible method for quantitative microbiota profiling. *PLoS ONE.* 2020;15:e227285.
76. Liu H, Zeng X, Zhang G, Hou C, Li N, Yu H, et al. Maternal milk and fecal microbes guide the spatiotemporal development of mucosa-associated microbiota and barrier function in the porcine neonatal gut. *Bmc Biol.* 2019;17:106.
77. Gohl DM, Vangay P, Garbe J, MacLean A, Hauge A, Becker A, et al. Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. *Nat Biotechnol.* 2016;34:942–9.
78. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 2013;41:e1.
79. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol.* 2019;37:852–7.
80. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome.* 2018;6:226.
81. Hsieh TC, Ma KH, Chao A. iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods Ecol Evol.* 2016;7:1451–6.
82. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* 2019;20:257.
83. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* 2017;27:824–34.
84. Zhao Z, Baltar F, Herndl GJ. Linking extracellular enzymes to phylogeny indicates a predominantly particle-associated lifestyle of deep-sea prokaryotes. *Sci Adv.* 2020;6:24354.

85. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *Bmc Bioinform.* 2010;11:119.
86. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinform.* 2012;28:3150–2.
87. Clausen PTL, Aarestrup FM, Lund O. Rapid and precise alignment of raw reads against redundant databases with KMA. *Bmc Bioinform.* 2018;19:307.
88. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 2018;46:W95–101.
89. Nissen JN, Johansen J, Allesøe RL, Sønderby CK, Armenteros JJA, Grønbech CH, et al. Improved metagenome binning and assembly using deep variational autoencoders. *Nat Biotechnol.* 2021;39:555–60.
90. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 2015;25:1043–55.
91. Stewart RD, Auffret MD, Roehe R, Watson M. Open prediction of polysaccharide utilisation loci (PUL) in 5414 public Bacteroidetes genomes using PULpy. 2018. <https://www.biorxiv.org/content/10.1101/421024v1.full>.
92. Hillmann B, Al-Ghalith GA, Shields-Cutler RR, Zhu Q, Gohl DM, Beckman KB, et al. Evaluating the information content of shallow shotgun metagenomics. *mSystems.* 2018;3:e00069–18.
93. Al-Ghalith GA, Hillmann B, Ang K, Shields-Cutler R, Knights D. SH17 is a self-learning pipeline for multipurpose short-read DNA quality control. *mSystems.* 2018;3:e00202.
94. McDonald JH. *Handbook of biological statistics*, vol. Baltimore, MD: Sparky House Publishing; 2009.
95. Bashan A, Gibson TE, Friedman J, Carey VJ, Weiss ST, Hohmann EL, et al. Universality of human microbial dynamics. *Nature.* 2016;534:259–62.
96. Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, et al. Stan: a probabilistic programming language. *Grantee Submission.* 2017;76:1–32.

ACKNOWLEDGEMENTS

We would like to thank Zepeng Qu, Dr. Chang Liu, and Prof. Shuang Jiang Liu for their help with the cultivation of gut bacteria strains. We would like to thank members of the LD lab for constructive comments on the manuscript. This research was supported by the National Key R&D Program of China (No. 2019YFA0906700, to LD), National Natural Science Foundation of China (Nos. 31971513, 32061143023, to LD), and China Postdoctoral Science Foundation (2020M682968, to HL).

AUTHOR CONTRIBUTIONS

HL and LD conceived the study. HL performed the experiments and analyzed the data. C. Liao analyzed the sequencing data and performed the computational analysis. LW, JT, JC, C. Lei and LZ assisted in experiments and/or data analysis. C. Liao, HL and LD wrote the manuscript with contributions from all coauthors.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41396-022-01253-4>.

Correspondence and requests for materials should be addressed to Lei Dai.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.