

MIT Open Access Articles

Predictability and persistence of prebiotic dietary supplementation in a healthy human cohort

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Gurry, Thomas et al. "Predictability and Persistence of Prebiotic Dietary Supplementation in a Healthy Human Cohort." *Scientific Reports* 8, 1 (August 2018): 12699 © 2018 The Author(s)

As Published: <http://dx.doi.org/10.1038/s41598-018-30783-1>

Publisher: Nature Publishing Group

Persistent URL: <http://hdl.handle.net/1721.1/120787>

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of use: Creative Commons Attribution 4.0 International license



SCIENTIFIC REPORTS



OPEN

Predictability and persistence of prebiotic dietary supplementation in a healthy human cohort

Thomas Gurry^{1,2,3}, HST Microbiome Consortium*, Sean M. Gibbons^{1,2,3},
Le Thanh Tu Nguyen^{1,2}, Sean M. Kearney^{1,2}, Ashwin Ananthakrishnan⁴, Xiaofang Jiang^{1,2,3},
Claire Duvallet^{1,2}, Zain Kassam^{1,5} & Eric J. Alm^{1,2,3}

Dietary interventions to manipulate the human gut microbiome for improved health have received increasing attention. However, their design has been limited by a lack of understanding of the quantitative impact of diet on a host's microbiota. We present a highly controlled diet perturbation experiment in a healthy, human cohort in which individual micronutrients are spiked in against a standardized background. We identify strong and predictable responses of specific microbes across participants consuming prebiotic spike-ins, at the level of both strains and functional genes, suggesting fine-scale resource partitioning in the human gut. No predictable responses to non-prebiotic micronutrients were found. Surprisingly, we did not observe decreases in day-to-day variability of the microbiota compared to a complex, varying diet, and instead found evidence of diet-induced stress and an associated loss of biodiversity. Our data offer insights into the effect of a low complexity diet on the gut microbiome, and suggest that effective personalized dietary interventions will rely on functional, strain-level characterization of a patient's microbiota.

A quantitative understanding of the forces that shape the composition of the microbiota *in vivo* is a prerequisite to rationally engineering or manipulating it toward a favorable clinical outcome. Diet has been shown to play a pivotal role in shaping the microbiota from an early age. For example, a study of the fecal microbiota of children from a rural African village in Burkina Faso found significant enrichment in Bacteroidetes and depletion in Firmicutes compared to Western counterparts, which is associated with a diet enriched in complex polysaccharides¹. In the same vein, a dietary exchange experiment between a rural African cohort and an urban, African-American cohort, led to concomitant changes in their microbiota and disease-associated molecular biomarkers for the duration of the exchange². This indicates that commensal microbiota may reflect long-term dietary trends but remain labile in the face of dietary perturbations. Indeed, it was also shown that short-term, large-scale changes in diet can result in rapid and reproducible effects on subjects' gut microbiome composition³. Observations such as these highlight the interplay between a subject's diet and their microbiota, and suggest that manipulating diet to engineer the gut microbiome may be a promising clinical intervention, together with rationally designed probiotics.

The interaction between short-term diet and the host microbiome is also an important and understudied confounder in many microbiome disease association studies. For example, both long-term dietary differences⁴ and alterations of the microbiota⁵ have independently been shown to be associated with differential rates of colon cancer; in the absence of a defined mechanism, it is unclear whether disease-associated microbial composition results from covariance with diet, which itself is causative, or whether the two factors are orthogonal and causally additive. A detailed and quantitative understanding of the impact of diet on the microbiome is therefore needed if the clinical vision of manipulating a patient's diet to engineer their gut microbiome's composition and/or metabolic output is to be achieved. However, identifying microbiome-diet interactions is complicated by a number of factors. Matching diets across patients is difficult, even if fixed meal plans are used, because participants may

¹Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA. ²Center for Microbiome Informatics and Therapeutics, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA.

³The Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA. ⁴Division of Gastroenterology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, 02114, USA. ⁵OpenBiome, Somerville, MA, 02143, USA.

*A comprehensive list of consortium members appears at the end of the paper. Correspondence and requests for materials should be addressed to E.J.A. (email: ejalm@mit.edu)

Arm	Number enrolled	Age (mean, range)	Number of bowel movements (preB, day 6)	Average number of Ensure calories consumed (day 3, day 6)	BMI (mean, range)	Gender distribution (females, males)
Inulin	11	27.4 (23–30)	1.0, 1.4	2156, 1975	23.6 (19–30)	5, 6
Pectin	11	24.8 (23–28)	1.3, 1.4	1728, 1920	23.5 (22–28)	5, 6
Protein	5	25.2 (22–26)	1.3, 1.2	1815, 1870	21.4 (18–25)	2, 3
Saturated fat	7	24.4 (23–31)	1.7, 1.6	2127, 2420	24.2 (19–27)	2, 5
Unsaturated fat	6	23.3 (23–33)	2.0, 1.2	1980, 2640	23.7 (20–29)	2, 4
Cellulose	10	29.0 (27–31)	1.3, 1.5	1886, 1860	24.1 (20–30)	5, 5
Control	10	25.4 (22–32)	0.8, 1.3	1760, 2310	22.4 (20–24)	4, 6

Table 1. Demographic and baseline data for each spike-in arm.

eat different quantities of individual components, altering the ratio of nutrients. Moreover, complex diets make it difficult to identify the role of individual micronutrients.

Previous studies have shown that microbiome composition responds to perturbations in diet^{3,6,7}. Indeed, some studies have shown that prebiotics affect the composition of the microbiota in predictable manner: for example, it has been shown that supplementation with inulin results increased abundance of *Bifidobacterium* and *Faecalibacterium* in a cohort of obese women⁸, as well as *Lactobacillus* and *Bifidobacterium* in rats⁹. A more recent double-blind, randomized, cross-over intervention study identified inulin-specific responses from *Anaerostipes*, *Bilophila* and *Bifidobacterium*⁷. However, perturbation studies in humans typically involve broad changes across complex mixtures of micronutrients³, measure the response to a particular spike-in against a variable diet background⁸, or perform relative abundance calculations from culture-based or qPCR methods^{10,11}. Resource partitioning in environmental bacteria has been shown to occur at exquisitely fine scales¹², yet little is known about resource partitioning in the human gut. If most species are generalists with respect to substrate use, then even a targeted addition can induce changes in many species, each competing for the same substrate. If most species are specialists, then we expect a strong response in a select few taxa.

Studies aimed at understanding how the microbiome will respond to a targeted change in a specific micronutrient remain logistically daunting. To overcome these challenges, we conducted a highly controlled feeding study and dietary perturbation experiment in a healthy human cohort, in which participants were placed on a standardized liquid nutritional meal-replacement for six days. Against this controlled dietary background, we investigated the effect of individual micronutrient spike-ins, including several prebiotic supplements, to identify the microbial responders and assess the extent to which prebiotics result in predictable compositional changes in the human microbiome.

Results

Study design. To examine the response of the healthy human microbiome to targeted addition of micronutrients, we enrolled 60 healthy participants and randomized each to one of seven dietary spike-in arms using a simple randomization procedure: pectin, inulin, cellulose, fish oil (unsaturated fat), coconut oil (saturated fat), protein powder, and control (no spike-in). These spike-ins were chosen to cover the dominant nutrient categories typically present in a human diet, namely soluble fiber (inulin and pectin), insoluble fiber (cellulose), unsaturated fat (fish oil), saturated fat (coconut oil), and protein. Basic demographic data and the results of the randomization process are available in Table 1. In order to assess day-to-day variability within and between participants under their habitual, variable diet, we first collected two consecutive baseline stool samples, after which participants underwent a bowel cleanse using an over-the-counter osmotic laxative to remove remnants of their previous diet from their colon. The following day, they began the standardized diet period: for the first three days, participants consumed only a standardized liquid nutritional meal-replacement (Ensure Original, Abbot Nutrition) that is routinely used in both inpatient and outpatient care, and water. This was followed by three further days of liquid nutritional meal-replacement and water plus their assigned spike-in (Fig. 1a). The duration of three days for the initial equilibration period was chosen based on previous observations of autocorrelation dynamics in long-term time series of the gut microbiota in healthy humans, which found that most autocorrelations decay within 3 days¹³. Participants were instructed to consume liquid nutritional meal-replacement *ad libitum* to allow for differential caloric requirements without compromising the relative composition of the standardized diet, and requested to submit stool samples preserved in nucleic acid stabilization buffer upon natural passage, with a maximum of one sample per day. Analysis of the average daily caloric intakes on days 3 and 6 did not show significant differences in caloric intake in a given arm. A post-intervention sample was taken on the first day of the resumption of routine, normal diet, as well as a follow-up sample one week later, to assess the persistence of any diet-induced compositional changes on the microbiota. All samples were processed within three days of receipt with only a single freeze-thaw cycle between sample processing and DNA extraction, library preparation and sequencing. Precise details of the standardized diet can be found in the Methods section.

Response to nutrient addition. Targeted nutrient additions induced a targeted microbial response, suggesting a high-level of resource partitioning within the human gut. On average, a typical participant had 9 OTUs that exhibited a greater than tenfold cumulative gain in relative abundance over the spike-in period (including sample 'post1'), and 14 OTUs with a fivefold gain. Some, but not all, nutrients yielded specific responses consistent across multiple individuals. To do this, we used differential abundance tests between day 3 and day 6 in

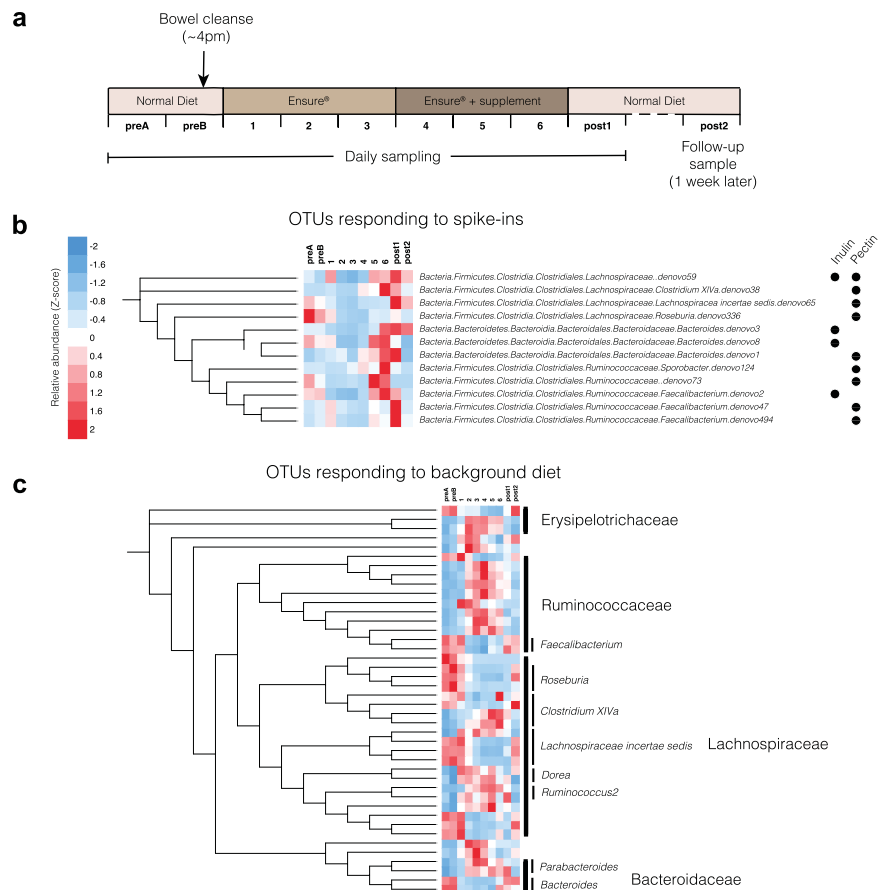


Figure 1. Nutritional meal replacement and prebiotic spike-ins result in reproducible changes across participants. **(a)** Schematic outlining the dietary and sampling regimen for the study. **(b)** OTUs that showed statistically significant (DESeq. 2, FDR < 0.1) differential abundance on day 6 compared to day 3 in response to particular spike-ins. Mean relative abundances are computed across all participants and then converted to Z-scores across timepoints, to illustrate relative changes through time. **(c)** OTUs that showed statistically significant (Wilcoxon Rank-Sum test, FDR < 0.1; $N = 39$) differential abundance on day 3 compared to baseline (preB). Complete RDP taxonomies can be found in Fig. S1.

each arm using DESeq2¹⁴ using raw 16S rRNA counts data (FDR < 0.1). DESeq. 2 has been shown to have high sensitivity on small sample numbers ($N < 20$)¹⁵. The only spike-ins with predictable and statistically significant responses across participants were pectin and inulin (Fig. 1b). This underscores their identity as prebiotics capable of altering microbial composition, as the other nutrients considered are typically absorbed in the small intestine and therefore not as likely to directly affect the composition of the colonic and stool microbiota. OTUs that responded to pectin or inulin generally reached their peak between day 5 and the first day of resuming a normal diet (post1). This is consistent with typical transit times of food in the gut, which vary by person but have been previously shown to involve a period of approximately 24–48 hours by using food dyes³. Responders to the prebiotic spike-ins were limited to the Clostridial *Lachnospiraceae* and *Ruminococcaceae* families, and the *Bacteroides* genus. While taxonomies were assigned using the RDP classifier¹⁶ and therefore did not include species-level annotation, we performed BLAST searches of the centroid 16S rRNA sequences for OTUs of interest against the NR database on NCBI. Notably, members of the same species appeared to respond to different spike-ins (Fig. 1b): in particular, the 16S rRNA sequences of the *Faecalibacterium* OTU responders, one of which responded to inulin and two to pectin, all mapped with 100% identity to *Faecalibacterium prausnitzii* genomes, suggesting that different commensal strains of the same species have different carbohydrate-active enzyme specificities. This is consistent with previous studies that reported *F. prausnitzii* blooms *in vivo* in participants given inulin as a supplement¹⁷, and also that commensal *F. prausnitzii* isolates can be cultured *in vitro* on pectin¹⁸.

Importantly, certain OTU responders exhibited highly predictable and reproducible behavior across people. In particular, a *Bacteroides* OTU that mapped with 100% identity to *B. uniformis* genomes on the NR database bloomed strongly in inulin-consuming participants, reaching relative abundances between four- and sixteen-fold greater than that on day 3 in the same participant after consuming the spike-in (Fig. 2a). In contrast, a number of OTUs appeared to respond to specific spike-ins in some participants but not others. These were undetectable using a DESeq. 2-based differential abundance test across all participants in the given arm, but were identifiable through an analysis of the individual trajectories between days 3 and post1. In particular, the cellulose arm was

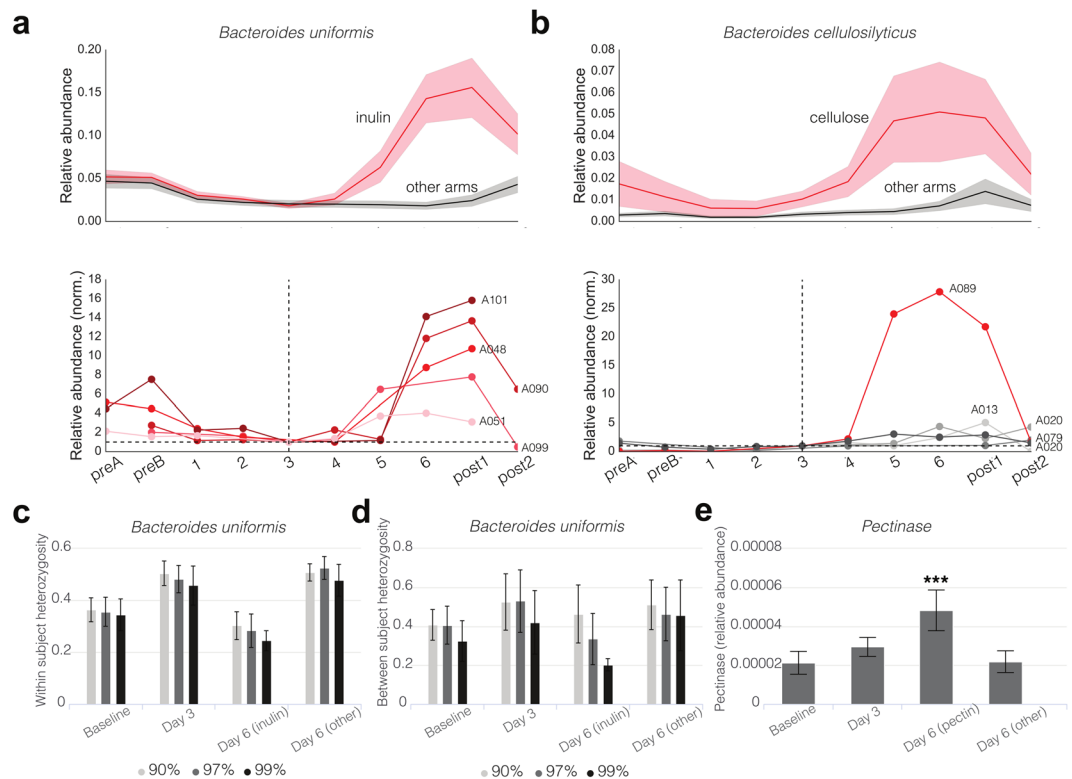


Figure 2. Inulin and pectin result in the enrichment of specific strains and specific enzymes in the metagenome across participants; in contrast, the same nutrient spike-in can also result in participant-specific responses. **(a)** *Bacteroides uniformis* blooms in response to inulin supplementation across all participants. The top plot shows mean and standard deviations of relative abundances across participants, and the bottom shows individual participant timeseries, normalized to relative abundance on day 3 for comparison purposes. **(b)** Same visualization for *Bacteroides cellulosilyticus*. **(c)** Average within-participant SNP heterozygosity in AMPHORA genes of *B. uniformis*. **(d)** Average between-participant SNP heterozygosity in AMPHORA genes of *B. uniformis*. **(e)** Relative abundance of pectinase enzymes in the metagenome. Statistical significance of the difference in abundance of pectinases on day 6 in the pectin arm was computed by a Wilcoxon Rank-Sum test compared to the abundance of pectinases on day 6 in the other arms ($p < 10^{-5}$).

significantly enriched with these person-specific responses. These responses can take two forms: in the first kind, the bacterial responder was present in most or all participants, but only bloomed in a subset of them, while in the second kind, the responder was not present in all participants but when it was, bloomed reproducibly. The most notable example of the first kind was an OTU which mapped with 100% identity to *Bacteroides cellulosilyticus*, a known cellulose degrader and commensal inhabitant of the human gut¹⁹, which exhibited an extremely strong response in a single person, despite being present in all other participants within that arm (Fig. 2b). In contrast, Archaeal methanogens were only present in a subset of people, but in those participants bloomed in response to cellulose (Fig. S2). This finding is consistent with previous observations that only a subset of humans are known to harbor commensal Methanobacteria, which utilize hydrogen produced by certain bacterial cellulose degraders for methanogenesis^{20,21}. It is also relevant in the context of reducing the net microbial production of hydrogen (as opposed to methane), which is one of the bases for a low FODMAP diet in the context of treating Irritable Bowel Syndrome²².

Strain-level dynamics. It is now known that polysaccharide specificity occurs at the species level²³. For example, species within the *Bacteroides* exhibit remarkable diversity in polysaccharide utilization^{24,25}. However, it is unclear to what extent individual strains exhibit differential polysaccharide utilization phenotypes. To investigate this point, we used shotgun metagenomics sequencing data to compute the mean heterozygosity of Single Nucleotide Polymorphisms (SNPs) within different species as a proxy for strain-level diversity (cf. Methods). The only bacterium that exhibited statistically significant differences in heterozygosity on day 6 compared to day 3 was *B. uniformis*, which had substantially lower heterozygosity in participants in the inulin arm on day 6 (Fig. 2c). This signal was retained (and indeed, stronger) as the identity cut-off for read mapping was increased up to 99%. This lower heterozygosity combined with increased relative abundance of the species in participants receiving inulin supplementation suggested that a specific strain of *B. uniformis* was being enriched in response to inulin.

Upon performing the same heterozygosity computation for *B. uniformis*, but this time between participants (i.e. the probability that overlapping read pairs, one from each person, exhibited a different allele), we observed a similar decrease in heterozygosity when 97% and 99% identity cut-offs were used, suggesting the same strain was

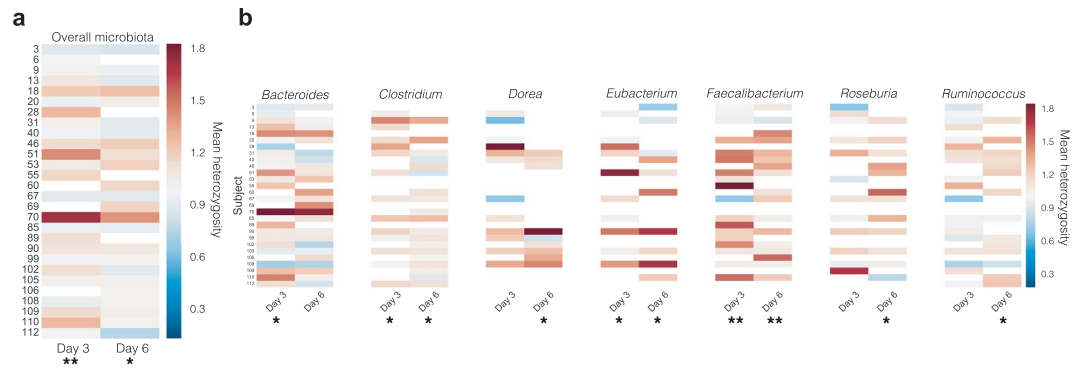


Figure 3. A diet consisting exclusively of nutritional meal replacement results in a flattening of the strain fitness landscape across the dominant genera of the gut microbiota. **(a)** Mean AMPHORA gene SNP heterozygosities within each participant on days 3 and 6, averaged across all bacterial species and normalized to baseline levels (preB). **(b)** Mean AMPHORA gene SNP heterozygosities averaged across the most abundant genera in the gut.

being enriched across different participants (Fig. 2d). These data indicate that different strains of *B. uniformis* have different inulin-utilization phenotypes.

Functional gene content. Since *B. uniformis* has a known inulin-specific polysaccharide utilization locus (PUL) in its genome²⁴, we reasoned that the enrichment of different OTUs or strains could be attributed to the presence of specific genomic carbohydrate active enzymes, and so mapped the metagenomic reads against HMM profiles for inulinases and pectinases in dbCAN²⁶. Consistent with this hypothesis, we observed an enrichment of pectinases in the metagenomes of participants in the pectin arm on day 6 (Fig. 2e). Surprisingly, however, we did not observe any such differences for inulinases in the inulin arm, perhaps due to poor annotation and characterization of these glycoside hydrolases and PULs in the human gut metagenome.

Response to background diet. Having shown that the response to a targeted dietary perturbation was both specific to the spike-in and in some cases generalizable across participants, we next turned our attention to the large-scale dietary perturbation built in to our study design: the switch to a standardized diet. We first looked for taxa that reproducibly responded to the diet change across participants by identifying OTUs that were differentially abundant on day 3 compared to baseline (preB) using a Wilcoxon Rank-Sum Test (FDR < 0.1, Benjamini/Hochberg procedure). OTUs that responded to the dietary perturbation generally fell into two clearly identifiable groups which followed two opposing dynamic trajectories (Figs 1c and S1): one group of OTUs increased in abundance on nutritional meal-replacement compared to baseline (preB), only to become depleted once a normal diet was resumed, and a second group of OTUs became depleted on nutritional meal-replacement relative to baseline (preB), only to return upon resumption of normal diet. The majority of OTU responders in both groups belonged to the Clostridial bacterial families *Ruminococcaceae* and *Lachnospiraceae*, with some exceptions (most notably *Erysipelotrichaceae*, *Deltaproteobacteria*, *Parabacteroides* and *Bacteroides*). Interestingly, there was no apparent high-level phylogenetic association with either group: indeed, OTUs assigned to the same bacterial genus by RDP sometimes exhibited opposing dynamics (e.g. OTUs in *Clostridium XIVa*).

Since the specific liquid nutritional meal-replacement used in this study is frequently used by physicians, dietitians and healthy individuals in a clinical context, our dataset offered a unique perspective on its effect on the microbiota of otherwise healthy humans. Aggregating participants from all spike-in arms, we counted the number of OTUs present and absent before and after the study, and found that participants lost an average of 32.3% of their total number of baseline (sample preB) 16S OTUs during the period of fixed diet, only 16.1% of which were regained after a week of resuming a normal diet. It is possible that the bowel cleanse was responsible for this loss, so to test the extent to which the cleanse might be able to explain the loss of diversity, we reasoned that organisms that were still present on day 2 of the liquid nutritional meal-replacement were unlikely to have been lost as direct result of the cleanse alone. Of the OTUs lost by the end of the prescribed diet, approximately half (17.6%) were lost by day 2, and the remainder was lost after day 2. Of course, even for the latter organisms, we cannot eliminate the possibility that the cleanse did not act in concert with the diet as a driving force for their disappearance. Nonetheless, these data suggest a loss of approximately 16% in the measurable number of OTUs. We then repeated the SNP-level heterozygosity analysis described above for all species in our reference genome set, and obtained mean heterozygosity values across all AMPHORA genes in all species for a given participant (Fig. 3a). We found that overall, the mean heterozygosity of SNPs in the entire microbiota increased significantly from baseline (preB) to day 3, a change that was maintained (although slightly diminished) by day 6. The effect for *Faecalibacterium*, an important commensal organism that has been negatively associated with Crohn's disease²⁷, was particularly strong (Fig. 3b). This relative increase in the strain-level diversity of the microbiota suggested a flattening of the fitness landscape between strains. This observation is reminiscent of previous work in *E. coli* that showed certain environmental stresses can alleviate the effect of deleterious mutations and therefore reduce fitness differences between strains²⁸.

We hypothesized that if the dietary change led to an increase in stress for the bacterial hosts, then we might expect to see an increase in active phage. The relative abundance of bacteriophage DNA in metagenomic

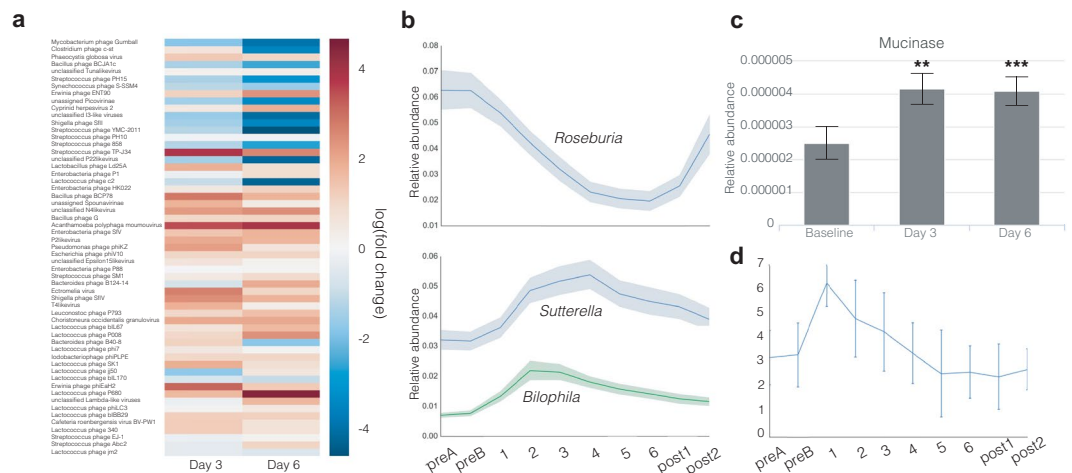


Figure 4. Nutritional meal replacement results in stress-like responses in the microbiota. **(a)** Log-transformed mean relative abundance of various bacteriophages in all participants, normalized to baseline levels (preB). **(b)** Mean and standard error timeseries of the relative abundance of *Roseburia*, *Bilophila* and *Sutterella*. **(c)** Relative abundance of mucinase genes on day preB, day 3 and day 6. **(d)** Bristol Stool Scale of all participants throughout the study.

sequences on days 3 and 6 relative to preB (Fig. 4a) suggested that many lysogenic phages were entering the lytic phase of their life cycles in response to the dietary perturbation.

The most striking and reproducible effect of the nutritional meal-replacement diet on microbiome composition was a decrease in the relative abundance of members of the *Roseburia* genus (Fig. 4b), an important and abundant butyrate-producing, mucosa-colonizing commensal that has been shown to be negatively associated with Crohn's disease and ulcerative colitis, types of inflammatory bowel disease (IBD)^{29,30}, as well as positively associated with obesity³¹. Concomitantly, a transient increase in Proteobacteria was observed: specifically, the genera *Bilophila* and *Sutterella* were found to transiently increase in relative abundance during the diet, peaking on days 2 and 4, respectively, before slowly returning to their baseline (preB) levels. While the physiological effect of these changes on the host is difficult to assess from genomic data alone, it is worth noting that both of these Proteobacterial genera have potential associations with disease: *Bilophila* has been associated with the production of sulfide, which leads to degradation of host mucin and induces colitis in mouse models^{32,33}, while *Sutterella* has been associated with insulin resistance in obese patients, and autism with gastrointestinal disturbance in children^{34,35}.

At the functional level, we observed significant increases in the relative abundance of mucinase genes in participants' metagenome in on days 3 and 6 relative to preB (Fig. 4c), consistent with a diet significantly impoverished in dietary fibers and complex polysaccharides, in which the colonic microbiota rely on host mucins as a carbon source³⁶. As in the mouse models that first documented this observation, the effect was not rescued by any of the individual prebiotic spike-ins³⁶. The increase in mucinases may also be linked to the depletion of *Roseburia* noted earlier: as a mucosal inhabitant, it is likely that destruction of the host mucin disrupted its native ecological niche. Importantly, loss of *Roseburia* was partially reversed one week after resuming normal diet (Fig. 4b), suggesting that the resumption of a more complex diet for a period of one week was able to partially restore the mucosal layer and its *Roseburia* inhabitants.

To quantify the extent of change induced by diet as a function of time, we computed a weighted Unifrac distance between participant pairs before and after the prescribed diet. We found that participants resembled each other more closely on day 6 than they did at baseline (preB), but only marginally so. However, the effect was stronger in the prebiotic spike-in arms (PERMANOVA, $p = 0.04$), suggesting that prebiotics led to a more predictable and reproducible response across participants compared to the standardized nutritional meal-replacement diet (Fig. 5a). We reasoned that this could be explained by the chemical nature of the diet: in the case of prebiotics, only a small set of organisms appeared to respond, but their responses were consistent across people (Figs 1b and 2a); in contrast, though the chemical composition of the standardized liquid diet was identical in all participants in the study, it was dominated by processed sugars, which are primarily absorbed in the small intestine, whose remainders act as ubiquitous substrates that would be chemically accessible to a larger fraction of the microbiota, thereby increasing stochasticity. We hypothesize that large inputs of non-specific substrates affect each participant's community differently and as a function of their initial composition.

Previous studies showed extensive day-to-day variation in the gut microbiome³⁷, hypothesizing that a large fraction of this variance may be due to diet¹³. We asked whether keeping diet standardized would reduce daily variation. We computed Weighted Unifrac distances between adjacent days within a person during a period of variable diet (samples 'preA' and 'preB') and compared this with Weighted Unifrac distances within the same participant on adjacent days in which the diet was identical (days 5 and 6). To our surprise, we found that there was no discernible increase in day-to-day similarity on a completely identical diet, irrespective of whether participants were given prebiotic spike-ins (Fig. 5a). These data suggest that short-term dietary fluctuations on daily

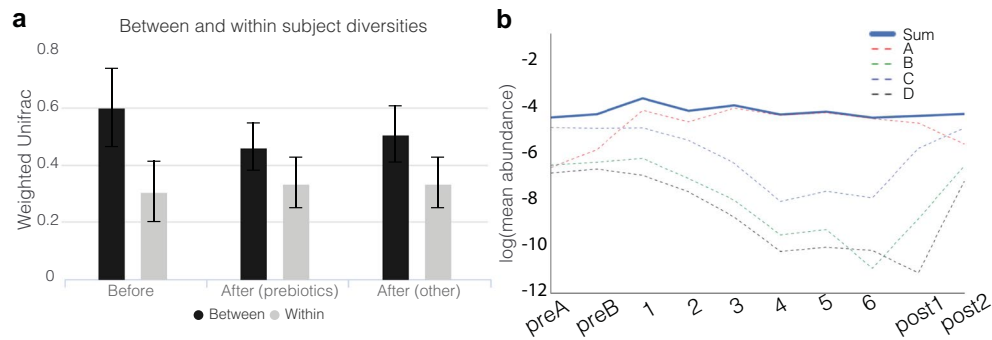


Figure 5. Participants' day-to-day variability in the composition of the microbiota does not decrease on identical diets, but between-participant similarity increases. **(a)** Weighted Unifrac values between and within participants on day preB (before) and day 6 (after). Error bars represent standard deviations. **(b)** Log-transformed mean relative abundance of the four distinct OTUs given a taxonomic assignment of *Lachnospiraceae incertae sedis* at the genus level and that were found to be differentially abundant on day 3 compared to baseline (preB). Their combined relative abundance is shown in a solid line. For clarity, the *denovo* OTU IDs are labeled as follows: *denovo6* - A; *denovo84* - B; *denovo65* - C; *denovo155* - D.

timescales contribute in an only very limited manner to the day-to-day variability observed in a participant's fecal microbiota, and that other variables, including fluctuations in host lifestyle, physiology, immunology, and perhaps even experimental noise are more important factors.

We found that a participant's sample could reliably be classified as pre- or post-nutritional meal-replacement. Two Random Forest Classifiers (RFCs) were constructed to classify samples as either baseline (preB) or the first day of resuming a normal diet (post1), and baseline (preB) or one week after resuming a normal diet (post2), respectively. In the first classifier, we obtained an AUC of 0.93 ($p < 10^{-7}$), while in the second, we obtained an AUC of 0.90 ($p < 10^{-5}$) (Fig. S3). Both classifiers had a classification accuracy (fraction of correctly classified cases during cross-validation) of 83%. These data indicate that there remain effects of the fixed diet on the overall microbiota even one week after resuming a normal diet (Table S2). The top 10 features of the preB/post1 RFC included three of the four OTUs assigned to *Lachnospiraceae incertae sedis* that appeared in Fig. 1c. Digging deeper, we found that these OTUs were indeed depleted on nutritional meal-replacement, while the remaining OTU replaces them, resulting in a constant combined relative abundance through time (Fig. 5b). These dynamics suggest a switch in occupancy of relative organisms in a particular niche in response to the dietary perturbation.

Discussion

When designing our study, we hypothesized that by introducing a constant diet background, we would improve our ability to detect microbial responders to particular prebiotic spike-ins. Indeed, we were able to detect reproducible microbial responses to both inulin and pectin across participants in relatively small numbers for each arm (<10) after filtering for early withdrawals or insufficient sampling. These responses were consistent across participants, including at the level of strains in the case of *B. uniformis*, indicating a strong predictability of the effects of prebiotic supplementation on the microbiota.

Interestingly, we did not observe significant blooms of *Bifidobacterium* in response to inulin, which have previously been documented in the literature^{10,11,17,38}. While we cannot conclusively explain this discrepancy, it is worth noting that many such studies are culture-based, emphasizing the need for further study and verification of these phenomena using modern sequencing technologies directly applied to stool. It is also worth noting that the three *Bifidobacterium* OTUs present in the dataset were only present in a subset of the participants in the inulin arm, which necessarily reduced our statistical power to detect responses thereof. In the case of at least one such non-culture-based analysis of the Bifidogenic nature of inulin, subjects were specifically chosen to contain appreciable amounts of *Bifidobacteria*¹⁷. This highlights the difficulty in observing statistically significant, reproducible responses of the microbiota *in vivo* in a cohort containing randomly chosen participants who may or may not contain the bacteria in question.

Moreover, in contrast to the predictability of inulin- and pectin-specific responses, certain responses were not detectable through traditional statistical means, because they only occurred in a subset of participants. The most noteworthy example was *B. cellulosilyticus*, which appeared to bloom strongly in response to cellulose in only one person. Archaeal methanogens also bloomed in responses to cellulose, but only in participants in which they were present. It is possible that these two cases are one and the same, i.e. that the *B. cellulosilyticus* OTU contains multiple strains with different functional properties, and the strain capable of degrading cellulose was present only in that one participant. Alternatively, it is possible that these person-specific blooms are manifestations of complex microbial networks on which this species depends to exhibit this phenotype. More broadly, these data offer a tantalizing glimpse into the potential for personalized nutritional interventions that account for strain-level composition of a patient's microbiota.

We also hypothesized that standardizing diet across participants would reduce the quantitative contribution of short-term dietary fluctuations on the microbiota, a concern in the design of clinical studies aiming to investigate the association between microbial composition or fluctuations and disease. However, after going to great lengths to place all participants on a standardized, liquid diet consisting of a nutritional meal-replacement, we

found that the diet introduced an axis of variation in the microbiota that did not appear to reach full stationarity after 6 days, and was further complicated by the difficulty of adhering to a liquid diet. From a practical standpoint, these data indicate that standardizing diet in a clinical cohort in this manner will have limited returns in reducing noise introduced by short-term dietary fluctuations, and may even introduce additional confounders. Averaging over additional participants or multiple timepoints may be more suitable means for reducing the impact of diet on other studies.

Moreover, the prescribed diet appeared to result in diet-induced stress on the participants' microbiota, with a flattening of the overall fitness landscape at the level of strains, an increase in the relative abundance of a large number of phages, and increases in mucinase abundance in the metagenome, indicative of increased consumption of host mucin as a carbon source and degradation of the mucosal layer by the microbiota. While the diet-induced stress is potentially confounded with stress induced from the bowel cleanse, the reported changes persisted from days 3 to 6, well after bowel movements returned to baseline consistency (Fig. 4d), indicating that the prescribed diet was a likely causative factor underlying these observations. While we did not experimentally measure changes to the mucosa, it is worth noting that thinning of the defensive mucosal layer is associated with a number of intestinal diseases, including colon cancer and IBD^{39–41}. It is likely that the chemical contents of the nutritional meal-replacement we used, which consist to a large extent of processed sugars easily converted to glucose, can explain these changes. The ubiquity of this substrate likely shrinks the phenotypic differences between strains, and favors fast growing organisms over the more slowly dividing host commensals, as suggested by the Proteobacterial responders (Fig. 4b). Taken together, these data suggest that clinical use of dietary supplementation or meal replacement strategies would benefit from more thoughtful formulations designed with the microbiota in mind, by including a higher content of dietary fibers and complex polysaccharides selected for their prebiotic qualities, in order to minimize potentially detrimental effects to the host.

Methods

Experimental model and participant details. 60 healthy human volunteers, consisting of 25 females and 35 males, were enrolled into the study under the supervision of the MIT Committee on the Use of Humans as Experimental Subjects (COUHES), who approved the study under protocol number 1504007066A002 (trial ID: ISRCTN53935058, registered 2/23/18). All participants provided written, informed consent, and the study was conducted in accordance with the relevant guidelines and regulations. To be included, participants had to be between 18 and 70 years of age and have a BMI between 18 and 30. Exclusion criteria were food allergies or dietary intolerances of any kind, a history of irritable bowel syndrome, inflammatory bowel disease, Type-2 diabetes, kidney disease or intestinal obstruction, allergy to polyethylene glycol, antibiotics treatment in the 6 months leading up to the study, untreated *in situ* colorectal cancer, or currently pregnant, planning to get pregnant in the next 60 days, or breast-feeding. Enrollment occurred between November 2015 and December 2016. Baseline cohort data and the results of the randomization process are available in Table 1.

Method details. *Study logistics.* Upon being consented into the study, participants were first randomized to a spike-in arm and were then provided with a kit consisting of materials for a bowel cleanse, sufficient stool collection hats and labeled Para-Pak vials (Meridian Biosciences, Inc.) with 5 ml RNALater (Ambion, Inc.) for daily sampling, in addition to instructions for appropriate sampling strategy, and three daily portions of their allocated spike-in. They were also provided with several locations in which stool samples could be dropped off and additional liquid meal-replacement could be picked up, and were instructed to collect samples whenever they naturally passed stool, with at most one sample per day. Samples were collected daily by study coordinators at all drop-off locations, and transported to a Biosafety Level 2 laboratory at MIT, where they were processed daily in batches, and no later than three days after passage. In addition, participants completed a daily questionnaire to record Bristol Stool Scale and bowel movement frequency, report any notable changes in health that did not relate to the study, including consuming antibiotics and medication, and inform study coordinators of lapses in adherence to the standardized diet, as well as document the amount of liquid meal-replacement consumed that day. Samples obtained after any reported lapses were excluded from the analysis.

Dietary regimen. All participants included in the analysis followed the dietary regimen outlined in Fig. 1a. For two days prior to beginning the study, participants were instructed to continue their habitual diet and provide two baseline stool samples (preA and preB). After obtaining the second stool sample, participants underwent a bowel cleanse in the evening by consuming 150 g of over-the-counter osmotic laxative PEG 3350 (MiraLAX, Bayer) and 64 oz bottles of electrolyte solution (Gatorade, PepsiCo) and were instructed to consume only clear liquids during the cleanse. Upon waking (day 1), participants began the standardized diet consisting entirely of liquid, nutritional meal-replacement (Ensure Original, Abbott Nutrition) and water for a period of six days. They were allowed to consume both *ad libitum* to avoid unnecessary weight loss and allow for heterogeneity in basal metabolic rates and daily caloric requirements. The liquid, nutritional meal-replacement was available in several flavors: vanilla, chocolate, strawberry, butter pecan, and coffee latte. Dark chocolate was excluded due to an increased amount of dietary fiber compared to the other flavors, which shared the same composition. Participants were allowed to choose freely between flavors, but were requested to maintain a consistent daily combination if they intended to consume several flavors. Nutritional composition of a single serving (one 8 oz bottle) of the liquid, nutritional meal replacement reported by Abbott Nutrition is as follows: 220 calories, consisting of 6 g total fat (1 g saturated fat, 0 g trans fat, 2 g polyunsaturated fat, 3 g monounsaturated fat), <5 mg Cholesterol, 200 mg Sodium, 370 mg Potassium, 32 g total carbohydrate (<1 g dietary fiber, 15 g sugars, derived mostly from corn maltodextrin and glucose), and 9 g protein. This consists of a diet that is significantly impoverished in dietary fiber, which we reasoned would improve our ability to detect changes in response to prebiotic spike-ins.

Participants were also provided with three daily doses of their allocated spike-in, to be consumed on days 4, 5 and 6. Spike-ins were given in pre-weighed powder form for inulin, pectin, cellulose and protein powder mixture, while fish oil was provided in pill form, and coconut oil was provided raw in pre-weighed tubes. Participants were instructed to consume their spike-ins throughout the day by mixing them either with a serving of liquid meal-replacement or water, and to record any lapses in either adherence to the standardized diet or to the consumption of their allocated spike-in on the daily questionnaires provided. The allocated daily doses of each spike-in were as follows: 10 g/day of inulin (CAS number 9005-80-5, Alfa Aesar); 35 g/day of pectin (CAS number 9000-69-5, MP Biomedicals, LLC); 20 g/day of cellulose (NutriCology); 40 g/day coconut oil (Kirkland Signature Organic Coconut Oil); 6 pills of fish oil each containing 1050 mg of omega-3, -5, -6, -7, -9 and -11 unsaturated fatty acids (Kirkland Signature Wild Alaskan Fish Oil); 51 g of protein powder, consisting of 25.5 g of whey protein (Optimum Nutrition Gold Standard 100% Whey Protein Powder) and 25.5 g of casein protein (Optimum Nutrition Gold Standard 100% Casein Protein Powder); no spike-in for control. Spike-in doses were set at 150% of the Recommended Daily Allowance (RDA) for that particular nutrient, where available. In the absence of a RDA (e.g. inulin, pectin, and cellulose), doses were set at the highest dose that the investigators determined could reasonably be ingested by a participant over the course of a day without experiencing discomfort.

Participants were instructed to consume the standardized diet plus allocated spike-in until the end of day 6. Two follow-up samples were also obtained, one on the day of resuming a normal, variable diet (post1) and another a week later (post2).

Stool sample processing and DNA sequencing. Participants collected approximately 1 g of stool and dissolved it in 5 ml RNALater (Ambion, Inc.) pre-aliquoted into Para-Pak vials (Meridian Biosciences, Inc.). In order to maximize our ability to obtain clean genomics signals and avoid experimental and sample processing confounders, great efforts were taken to ensure that all samples were treated in the same manner. Samples were picked up from daily drop-off locations by study coordinators daily, and washed with PBS the same day or no longer than 3 days upon passage before being frozen at -80°C . 29 of the 60 participants returned a complete set of daily stool samples for the entirety of the requested timeseries. Samples were aggregated through time and submitted together for DNA extraction, library preparation and DNA sequencing at the Genomics Platform at Broad Institute of MIT and Harvard. 16 S rRNA regions were amplified from all samples using a universal V4 primer. PreB, day 3 and day 6 samples were also sequenced using shotgun metagenomics sequencing. 16 S rRNA amplicon sequencing was performed on an Illumina MiSeq using v2 chemistry, and metagenomics sequencing was performed on an Illumina HiSeq. 2500 High Output flowcell using v4 chemistry.

Quantification and statistical analysis. *16S rRNA sequencing data analysis.* Raw read processing and OTU calling: Raw paired-end 16 S rRNA Illumina sequencing reads were merged with PEAR⁴², resulting in a total of 64,743,565 raw merged reads. Reads were then demultiplexed, and quality trimmed with a cut-off of $Q = 25$ using usearch⁴³, before being trimmed to a common length of 226 bases. Dereplicated reads were then clustered into OTUs to 97% identity using UPARSE⁴⁴. OTU centroids were assigned a taxonomy using the RDP classifier using an uncertainty cut-off of 0.5¹⁶. Since RDP does not return species annotation, we also performed BLAST searches of the centroid 16 S rRNA sequences for OTUs of interest against the NR database on NCBI⁴⁵. A species annotation was given if the top 10 alignments matched a species' genome with 100% identity. Samples with fewer than 5,000 reads (totaling 38 out of 501) were discarded from the analysis.

Phylogenetic trees and distance calculations: A phylogenetic tree was constructed from all 16 S OTU centroid sequences using the neighbor-joining based on uncorrected distances, implemented in the FastTree tool⁴⁶. Sub-trees in Fig. 1 were computed in the same manner but using only the OTU sequences of interest, and plotted using iTOL⁴⁷. Beta diversity calculations between samples (Fig. 5a) were computed from rarefied OTU tables with the Weighted Unifrac distance⁴⁸ using the scikit-bio Python package and the phylogenetic tree mentioned in the section above. Prior to computing beta diversity values, a rarefaction was performed on the OTU table by down-sampling each sample to 5,516 reads (the minimum read count in the dataset after filtering), to remove artifactual effects of differential read counts across samples. Statistical significance of the differences in between-participant beta diversities on day 6 in the prebiotic versus non-prebiotic arms was assessed with a PERMANOVA test ($N = 29$, 10,000 permutations) using the scikit-bio Python package.

Statistical tests of differential abundance: Statistical tests to identify OTU responders to spike-ins (Fig. 1b) were performed using DESeq. 2¹⁴. Participants who had both day 3 and day 6 samples were aggregated into their respective spike-in arms, and the test was performed between those two days. Due to the effects of attrition on participant numbers and since not all participants produced daily stool, day 5 samples were also included in the day 6 category in the absence of a day 6 sample. The number of participants in each arm that had the necessary samples to be included in the statistical tests was $N = 9$ pectin, $N = 5$ inulin, $N = 4$ cellulose, $N = 6$ in the saturated fat, $N = 3$ unsaturated fat, $N = 2$ protein, and $N = 6$ control. OTU responders were identified from each arm as OTUs whose FDR corrected q-value was below 0.1 in that arm and above 0.1 in the control arm.

OTU responders to the standardized, nutritional meal-replacement background (Figs 1c and S1) were identified using a Wilcoxon rank-sum test between preB and day 3 ($N = 39$), with a Benjamini/Hochberg FDR cut-off of 0.1.

Metagenomics sequencing data analysis. Raw read processing: Human reads were removed from raw forward and reverse Illumina Whole Genome Shotgun (WGS) reads by aligning to the hg18 human genome from the UCSC Genome Browser⁴⁹ using BWA⁵⁰. To remove biases introduced by PCR replicates, remaining reads were then dereplicated using PRINSEQ⁵¹, for all exact matches and 5' duplicates (prinseq option: -derep 2). Unique reads were then quality-trimmed with trimmomatic to $Q = 20$ ⁵², to a final total readcount of 1,695,900,645, or an

average sequencing depth of 15,007,970 reads per sample, with 15 out of 113 samples having fewer than 1,000,000 reads.

Carbohydrate active enzyme annotations: Functional annotations for reads were obtained from the CAZY database using dbCAN²⁶. Pre-processed metagenomics reads were converted to the 6 possible Open Reading Frames using EMBOSS Transeq⁵³, and aligned to the HMM profiles using HMMER3.0⁵⁴. The profiles searched were: GH32, GH91 and CBM38 for inulinases/fructanases, and CE8, CE13, PL1 and GT47 for pectinases. Only CE8 (shown in Fig. 2e) showed statistically significant differences as measured by Wilcoxon rank-sum tests on the relative abundances on days 3 and day 6. We also searched PL8 for hyaluronate lyases (EC 4.2.2.1), also known as ‘mucinases’ (as shown in Fig. 3). A 1e-3 E-value cutoff was used for alignments less than 80 amino acids long, and 1e-5 for longer alignments.

Single nucleotide polymorphism (SNP) heterozygosity calculations: Heterozygosity values reported in Figs 2 and 3 were computed using AMPHORA⁵⁵ to identify thirty-one single-copy phylogenetic markers in a set of 649 non-redundant reference genomes from the Human Microbiome Project⁵⁶. Metagenomics reads were mapped from each sample to these genes using BWA-mem⁵⁰ with the “-a” flag and a conservative cut-off of 90% sequence identity. The 90% identity cut-off was chosen based on previous work showing that a 97% distance in 16S rRNA, typically associated with the species barrier, is associated with a median distance of ~85% in AMPHORA genes⁵⁷. We also performed these alignments with cut-offs of 97% and 99%. These thirty-one alignments were then filtered for SNPs by removing all monomorphic sites. Next, all samples with zero coverage at more than 25% of the identified SNP loci were removed, and finally alignment sites with atypical coverage, defined as being greater than 1.5 standard deviations away from the mean coverage, were also removed. This left us with a set of informative single nucleotide polymorphic sites for each gene alignment. Within-participant heterozygosity values were computed at each site by considering the set of all reads in a sample that overlapped with that site, and calculating the probability that any of these two reads had the same allele. From the Hardy-Weinberg equilibrium, 1 minus this value gives the heterozygosity. Between-participant heterozygosity values for a given site were computed by considering the set of all reads overlapping with that site in two participants, and calculating the probability that two reads, one from each participant, share the same allele. Python code for performing these calculations from sorted BAM files can be found at the following Github link: <https://github.com/thomasgurry/strains/blob/master/heterozygosity.py>

Bacteriophage annotations: Relative abundance of phages were estimated using Kraken and the full NCBI phage database⁵⁸. Filtered read counts were converted to relative abundances by dividing by the total number of reads.

Machine learning. Random Forest Classifiers (RFCs) were built using the scikit-learn Python package. Classifiers were built using participants which contained one of each sample of the classification classes: for example, a participant had to contain one of ‘preA’ or ‘preB’ samples and a ‘post1’ sample to be included in the preB/post1 RFC. This ensured that each class was balanced and that no biases could be introduced from inclusion of samples from a participant in only one of the two classes. The preB/post1 and preB/post2 RFCs had $N = 26$ and $N = 23$ in each class, respectively. Cross-validation was performed using a ‘leave a participant out’ approach, i.e. by successively removing both samples from a participant samples and using them as test sets, in order to construct an average Receiver Operator Characteristic (ROC) from which to compute an AUC. *P*-values for each classifier were computed using Fisher exact tests on the resulting confusion matrices. Classification accuracy was computed from the confusion matrices as the fraction of correctly classified cases, i.e. $(TP + TN) / (TP + TN + FP + FN)$, where *TP*, *TN*, *FP* and *FN* are the number of true positives, true negatives, false positives and false negatives, respectively.

Data availability. The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References

- De Filippo, C. *et al.* Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc. Natl. Acad. Sci. USA* **107**, 14691–14696 (2010).
- O’Keefe, S. J. D. *et al.* Fat, fibre and cancer risk in African Americans and rural Africans. *Nat. Commun.* **6**, 6342 (2015).
- David, L. A. *et al.* Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559–563 (2014).
- Boyle, P. & Leon, M. E. Epidemiology of colorectal cancer. *Br. Med. Bull.* **64**, 1–25 (2002).
- Baxter, N. T., Ruffin, M. T., Rogers, M. A. M. & Schloss, P. D. Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions. *Genome Med.* **8**, 37 (2016).
- Minot, S. *et al.* The human gut virome: Inter-individual variation and dynamic response to diet. *Genome Res.* **21**, 1616–1625 (2011).
- Vandeputte, D. *et al.* Prebiotic inulin-type fructans induce specific changes in the human gut microbiota. *Gut* **66**, 1968–1974 (2017).
- Dewulf, E. M. *et al.* Insight into the prebiotic concept: lessons from an exploratory, double blind intervention study with inulin-type fructans in obese women. *Gut* [gutjnl-2012-303304](https://doi.org/10.1136/gutjnl-2012-303304), <https://doi.org/10.1136/gutjnl-2012-303304> (2012).
- Hoentjen, F. *et al.* Reduction of colitis by prebiotics in HLA-B27 transgenic rats is associated with microflora changes and immunomodulation. *Inflamm. Bowel Dis.* **11**, 977–985 (2005).
- Gibson, G. R., Beatty, E. R., Wang, X. & Cummings, J. H. Selective stimulation of bifidobacteria in the human colon by oligofructose and inulin. *Gastroenterology* **108**, 975–982 (1995).
- Kruse, H.-P., Kleessen, B. & Blaut, M. Effects of inulin on faecal bifidobacteria in human subjects. *Br. J. Nutr.* **82**, 375–382 (1999).
- Hunt, D. E. *et al.* Resource Partitioning and Sympatric Differentiation Among Closely Related Bacterioplankton. *Science* **320**, 1081–1085 (2008).
- Gibbons, S. M., Kearney, S. M., Smillie, C. S. & Alm, E. J. Two dynamic regimes in the human gut microbiome. *PLOS Comput. Biol.* **13**, e1005364 (2017).
- Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq. 2. *Genome Biol.* **15**, 550 (2014).

15. Weiss, S. *et al.* Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* **5**, 27 (2017).
16. Cole, J. R. *et al.* The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucleic Acids Res.* **33**, D294–D296 (2005).
17. Ramirez-Farias, C. *et al.* Effect of inulin on the human gut microbiota: stimulation of *Bifidobacterium adolescentis* and *Faecalibacterium prausnitzii*. *Br. J. Nutr.* **101**, 541–550 (2009).
18. Lopez-Siles, M. *et al.* Cultured Representatives of Two Major Phylogroups of Human Colonic *Faecalibacterium prausnitzii* Can Utilize Pectin, Uronic Acids, and Host-Derived Substrates for Growth. *Appl. Environ. Microbiol.* **78**, 420–428 (2012).
19. Wu, M. *et al.* Genetic determinants of *in vivo* fitness and diet responsiveness in multiple human gut *Bacteroides*. *Science* **350**, aac5992 (2015).
20. Bäckhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I. Host-Bacterial Mutualism in the Human Intestine. *Science* **307**, 1915–1920 (2005).
21. Chassard, C., Delmas, E., Robert, C. & Bernalier-Donadille, A. Cellulose-degrading microbial community of the human gut varies according to the presence or absence of methanogens. *FEMS Microbiol. Ecol.* **74**, 205–213 (2010).
22. Gibson, P. R. & Shepherd, S. J. Evidence-based dietary management of functional gastrointestinal symptoms: The FODMAP approach. *J. Gastroenterol. Hepatol.* **25**, 252–258 (2010).
23. Chung, W. S. F. *et al.* Modulation of the human gut microbiota by dietary fibres occurs at the species level. *BMC Biol.* **14**, 3 (2016).
24. Sonnenburg, E. D. *et al.* Specificity of Polysaccharide Use in Intestinal *Bacteroides* Species Determines Diet-Induced Microbiota Alterations. *Cell* **141**, 1241–1252 (2010).
25. Xu, J. *et al.* Evolution of Symbiotic Bacteria in the Distal Human Intestine. *PLOS Biol.* **5**, e156 (2007).
26. Yin, Y. *et al.* dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **40**, W445–W451 (2012).
27. Sokol, H. *et al.* *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc. Natl. Acad. Sci.* **105**, 16731–16736 (2008).
28. Kishony, R. & Leibler, S. Environmental stresses can alleviate the average deleterious effect of mutations. *J. Biol.* **2**, 14 (2003).
29. Maslowski, K. M. & Mackay, C. R. Diet, gut microbiota and immune responses. *Nat. Immunol.* **12**, 5–9 (2011).
30. Willing, B. P. *et al.* A Pyrosequencing Study in Twins Shows That Gastrointestinal Microbial Profiles Vary With Inflammatory Bowel Disease Phenotypes. *Gastroenterology* **139**, 1844–1854.e1 (2010).
31. Duvallet, C., Gibbons, S. M., Gurry, T., Irizarry, R. A. & Alm, E. J. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nat. Commun.* **8**, 1784, <https://doi.org/10.1038/s41467-017-01973-8> (2017).
32. Devkota, S. *et al.* Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in *Il10^{-/-}* mice. *Nature* **487**, 104–108 (2012).
33. Ijssennagger, N., van der Meer, R. & van Mil, S. W. C. Sulfide as a Mucus Barrier-Breaker in Inflammatory Bowel Disease? *Trends Mol. Med.* **22**, 190–199 (2016).
34. Moreno-Indias, I. *et al.* Insulin resistance is associated with specific gut microbiota in appendix samples from morbidly obese patients. *Am. J. Transl. Res.* **8**, 5672–5684 (2016).
35. Williams, B. L., Hornig, M., Parekh, T. & Lipkin, W. I. Application of Novel PCR-Based Methods for Detection, Quantitation, and Phylogenetic Characterization of *Sutterella* Species in Intestinal Biopsy Samples from Children with Autism and Gastrointestinal Disturbances. *mBio* **3**, e00261–11 (2012).
36. Desai, M. S. *et al.* A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. *Cell* **167**, 1339–1353.e21 (2016).
37. David, L. A. *et al.* Host lifestyle affects human microbiota on daily timescales. *Genome Biol.* **15**, R89 (2014).
38. Rossi, M. *et al.* Fermentation of Fructooligosaccharides and Inulin by *Bifidobacteria*: a Comparative Study of Pure and Fecal Cultures. *Appl. Environ. Microbiol.* **71**, 6150–6158 (2005).
39. Brockhausen, I., Schutzbach, J. & Kuhns, W. Glycoproteins and Their Relationship to Human Disease. *Cells Tissues Organs* **161**, 36–78 (1998).
40. Corfield, A. P. *et al.* Glycosylation patterns of mucins in colonic disease. *Biochem. Soc. Trans.* **23**, 840–845 (1995).
41. Shirazi, T., Longman, R. J., Corfield, A. P. & Probert, C. S. J. Mucins and inflammatory bowel disease. *Postgrad. Med. J.* **76**, 473–478 (2000).
42. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd merger. *Bioinforma. Oxf. Engl.* **30**, 614–620 (2014).
43. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
44. Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**, 996–998 (2013).
45. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **44**, D7–D19 (2016).
46. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009).
47. Letunic, I. & Bork, P. Interactive tree of life (iTOL)v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–245 (2016).
48. Lozupone, C. & Knight, R. UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. *Appl. Environ. Microbiol.* **71**, 8228–8235 (2005).
49. Kent, W. J. *et al.* The Human Genome Browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
50. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinforma. Oxf. Engl.* **25**, 1754–1760 (2009).
51. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863–864 (2011).
52. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
53. Goujon, M. *et al.* A new bioinformatics analysis tools framework at EMBL–EBI. *Nucleic Acids Res.* **38**, W695–W699 (2010).
54. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29–W37 (2011).
55. Wu, M. & Eisen, J. A. A simple, fast, and accurate method of phylogenomic inference. *Genome Biol.* **9**, R151 (2008).
56. Consortium, T. H. M. P. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
57. Smillie, C. S. *et al.* Strain Tracking Reveals the Determinants of Bacterial Engraftment in the Human Gut Following Fecal Microbiota Transplantation. *Cell Host Microbe* **23**, 229–240.e5 (2018).
58. Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15**, R46 (2014).

Acknowledgements

The authors wish to thank Dr. Ashley Vargas from the NIH for preliminary feedback on the study design, Dr. Vicki Mountain and Paige Swanson for logistical support, Dr. Jason Zhang for helpful discussions, and the Rasmussen family foundation for funding. This project was in part supported by award Number 132GM007753 from the National Institute of General Medical Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

Author Contributions

T.G. designed the study, performed participant recruitment and sample processing, analyzed the data and wrote the manuscript. H.S.T.M.C., P.H.D., S.G.F., T.K.H., C.M.-T., K.O.-B., A.S., S.L.T., M.L.V., K.K.W., M.P.W., T.I.Z. designed the study, performed participant recruitment, and reviewed the manuscript. S.G. assisted in analyzing the data. L.T.T.N. assisted in study design and sample processing. S.M.K. assisted in data analysis. A.A. reviewed study design and acted as study physician. X.J. assisted in data analysis. Z.K. and C.D. assisted in study design and writing the manuscript. E.J.A. designed the study, analyzed the data and wrote the manuscript. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-30783-1>.

Competing Interests: Dr. Ashwin Ananthakrishnan served on scientific advisory boards for Takeda and Abbvie.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018

Consortia
HST Microbiome Consortium*

**Paul H. Dannenberg^{6,7}, Samuel G. Finlayson^{6,7}, Travis K. Hughes^{6,7}, Claudio Macias-Trevino^{6,7},
Kwadwo Owusu-Boaitey^{6,7}, Andre Shomorony^{6,7}, Suan Lian Tuang^{6,7}, Max L. Valenstein^{6,7},
Kathy K. Wang^{6,7}, Michael Pei-hong Wu^{6,7} & Travis I. Zack^{6,7}**

⁶Harvard-MIT Program, Health Sciences and Technology, Cambridge, MA, 02139, USA. ⁷Institute for Medical Engineering and Sciences, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA.