HHS Public Access

Author manuscript

Science. Author manuscript; available in PMC 2016 April 02.

Published in final edited form as:

Science. 2015 October 2; 350(6256): aac5992. doi:10.1126/science.aac5992.

Genetic determinants of *in vivo* fitness and diet responsiveness in multiple human gut *Bacteroides*

Meng Wu[1,2], **Nathan P. McNulty**[1], **Dmitry A. Rodionov**[3,4], **Matvei S. Khoroshkin**[4], **Nicholas W. Griffin**[1,2], **Jiye Cheng**[1,2], **Phil Latreille**[5], **Randall A. Kerstetter**[5], **Nicolas Terrapon**[6], **Bernard Henrissat**[6,7], **Andrei L. Osterman**[3], and **Jeffrey I. Gordon**[1,*]

[1]Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, MO 63108 USA

[2]Center for Gut Microbiome and Nutrition Research, Washington University School of Medicine, St. Louis, MO 63108 USA

[3]Infectious and Inflammatory Disease Center, Sanford-Burnham-Prebys Medical Discovery Institute, La Jolla, 92037 CA USA

[4]A.A.Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow 127994, Russia

[5]Monsanto Company, St. Louis, MO 63167 USA

[6]Architecture et Fonction des Macromolécules Biologiques, Centre National de la Recherche Scientifique & Aix-Marseille Université 13288 Marseille cedex 9, France

[7]Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia

Abstract

Libraries of tens of thousands of transposon mutants generated from each of four human gut *Bacteroides* strains, two representing the same species, were introduced simultaneously into gnotobiotic mice together with 11 other wild-type strains to generate a 15-member artificial human gut microbiota. Mice received one of two distinct diets monotonously, or both in ordered sequence. Quantifying the abundance of mutants in different diet contexts allowed gene-level characterization of fitness determinants, niche, stability and resilience, and yielded a prebiotic

*Corresponding author. jgordon@wustl.edu.

Data deposition:

INSeq, COPRO-Seq, and microbial RNASeq datasets have been deposited in the European Nucleotide Archive (ENA) under accession number PRJEB9434. The finished and annotated genome sequences of *B. thetaiotaomicron* 7330, *B. ovatus* ATCC 8483 and *B. cellulosilyticus* WH2 have been deposited in NCBI under accession number PRJNA289334.

Declarations:

J.I.G. is cofounder of Matatu Inc., a company characterizing the role of diet-by-microbiota interactions in animal health. Nathan McNulty is currently an employee of Matatu, Inc. Andrei Osterman is an Adjunct Vice President for Research for Buffalo BioLabs, LLC.

Supplementary Materials:
Materials and Methods
Supplemental Results
Figures S1–S17
Tables S1–S14
References (28–44)

(arabinoxylan) that allowed targeted manipulation of the community. The approach described is generalizable and should be useful for defining mechanisms critical for sustaining and/or approaches for deliberately reconfiguring the highly adaptive and durable relationship between the human gut microbiota and host in ways that promote wellness.

The human gut microbiota is highly diverse (1–3). A current view is that strains acquired by an individual early in life persist for decades and that strains are shared among family members. The microbiota can rapidly adapt to changing conditions but the degree to which given sets of strains share or compete for niche space in the gut ecosystem is poorly understood. Identification of the genetic factors that define an organism's niche is important for understanding the mechanisms that determine community assembly, community responses to and recovery after various perturbations, and the food webs that link microbes to one another and to their host. Discovery of these factors should spawn new approaches for intentional manipulation of the functional properties of the microbiota.

In this report we describe an approach for simultaneously identifying genetic determinants of fitness for multiple members of a defined artificial human gut microbiota installed in gnotobiotic mice fed distinct diets monotonously or in an ordered sequence. We focus on four human gut *Bacteroides* strains: *B. cellulosilyticus* WH2, *B. ovatus* ATCC 8483, *B. thetaiotaomicron* 7330 and *B. thetaiotaomicron* VPI-5482 [finished genomes for the first three strains were obtained by combining PacBio and Illumina sequencing [see SM Materials and Methods and (4) for details]. Culture-independent surveys of a number of human populations indicate that all three of these *Bacteroides* species are prominently represented in the guts of healthy individuals (1, 2). All four strains contain numerous genes involved in the recognition and processing of otherwise indigestible dietary polysaccharides (4–7; Table S1). We apply genome-wide transposon mutagenesis to these three prominent human gut-derived *Bacteroides* species, and the two strain representatives of one of them, and colonize singly housed, germ-free mice with all four mutant libraries together with a defined consortium of 11 other wild-type human gut bacterial species representing major phylogenetic lineages present in the microbiota. The microbiome of this 15-member community encodes key metabolic functions identified in anaerobic food webs, including the ability to process polysaccharides to oligosaccharides and simple sugars, and ferment amino acids (7, 8).

This approach not only allowed us to examine assembly of the 15-member artificial community, but to characterize responses to dietary perturbations and recovery after these perturbations (stability and resilience) at the community level, the individual species/strain level, as well as the gene level. We identify shared as well as species- and strain-specific genetic and metabolic features that impact fitness of *Bacteroides* in the gut environment.

Characterizing multiple transposon mutant libraries simultaneously *in vivo* (multitaxon INSeq)

Each library we generated from the four bacterial strains was composed of 87,000–167,000 isogenic transposon (Tn) mutants. Each mutant contained a single site of insertion of the Tn

element. 81.5–91.8% of the open reading frames (ORFs) in these four *Bacteroides* contained a Tn insertion, allowing us to conclude that disruption of these genes did not preclude growth in the rich medium used to construct the libraries (see Supplemental Results for an analysis of 'essential' genes not represented in the various mutant libraries, with a focus on those involved in carbohydrate, amino acid, and vitamin/cofactor biosynthesis/catabolism).

Each ORF was covered by an average of 11 to 26 Tn insertions (Table S2; fig. S1A–D). These Tn mutant libraries were introduced into germ-free mice together with 11 wild-type species that are common constituents of the adult gut microbiota (Fig. S2A). To permit simultaneous analysis of multiple mutant libraries in the same recipient gnotobiotic mouse, a *mariner* transposon delivery vector with MmeI sites positioned at each end of the Tn (9, 10) was modified so that it contained two taxon-specific barcodes (Fig. S2B). MmeI digestion of microbial DNA prepared from the gut contents or feces of recipient gnotobiotic mice cleaves genomic DNA at a site 20–21 bp distal to the restriction enzyme's recognition site. The site of Tn insertion and the relative abundance of each Tn mutant in the input libraries introduced into mice and in the 'output' communities recovered from these animals were defined using a multi-taxon insertion sequencing (INSeq) protocol. This protocol enables sequencing of the flanking genomic sequence in addition to the taxon-specific barcode positioned within the transposon (Fig. S2B). A software package (SM Methods) was used to assign sequencing reads based on the sample-specific barcode and the taxon-specific barcode. Control experiments, using 'mock communities' composed of the 15-member collection of cultured human gut bacteria with different proportions of the mutant libraries (ranging from 0.5–40% of the community) revealed that this protocol had high specificity (91.6±0.84% of reads mapped to sites of Tn insertion), high sensitivity (a mutant library representing as little as 0.5% of the entire community could be characterized), and good reproducibility (R^2>0.9; n=5 technical replicates/mock community) (Table S2, see SM Methods and SM Results for details).

To identify mutations that were deleterious in the gut, we calculated a "fitness index" for each gene based on the relative abundance of its INSeq reads in the output community (e.g., fecal microbiota) compared to input library. We observed that the output to input log-ratios for the entire mutant population were bi-modally distributed, with the vast majority of mutants (genes) falling in a normal distribution to the right of a much smaller distribution with lower fitness. To statistically distinguish these two populations, we implemented the Expectation Maximization (EM) algorithm to iteratively estimate the unknown parameters (mean and variance) for each population using a likelihood function based on observed data (i.e., the log-ratios for each individual gene) (see SM Methods for details). To identify deleterious mutations, we calculated a z-score (fitness index) for the difference between each gene's log-ratio and the mean of the normal distribution (Fig. S2C–E). A p-value was determined based on a normal distribution using a Z-test, and a q-value assigned after applying a multiple hypothesis testing correction (FDR). To avoid misidentifying genes as significant fitness determinants, we only considered those that were significantly depleted across biological replicates. *In silico* simulations revealed that this EM algorithm-based procedure yielded a false positive rate of less than 5% for mutants in genes whose relative abundance was >0.002% (Fig. S3B). These simulations were verified by applying the

analysis pipeline to our mock communities (Table S2). ORFs satisfying the criteria of having a significant *q*-value in one diet condition, and no fitness defects in any biological replicates under another diet condition, were defined as 'diet-specific fitness factors'.

Community assembly, stability and resilience in response to different diets

Adult (10 to 12 week-old) male germ-free C57BL/6J mice were placed on either a low-fat/high-plant polysaccharide (LF/HPP) chow or a diet where calories were largely derived from hydrogenated vegetable shortening and beef tallow, and where carbohydrates were limited to sucrose, corn starch, and maltodextrin (plus an indigestible cellulose binder) [high saturated fat/high simple sugar (HF/HS) diet]. Diets were introduced 7 d before animals received a single oral gavage of the artificial community containing the four mutagenized strain libraries and 11 wild-type strains (Fig. 1A). Control experiments used a 15-member community composed of all wildtype strains, or 15-member communities that only contained a single mutant library (Fig. S4A). Groups of singly caged mice were maintained on either the LF/HPP or HF/HS diet for the duration of the experiment, while other groups were subjected to a reciprocal set of diet oscillations (LF/HPP→HF/HS→LF/HPP or HF/HS→LF/HPP→HF/HS) (n=5 mice/diet treatment/community type/gnotobiotic isolator; fig. 1A). Short-read shotgun sequencing (COmmunity PROfiling by Sequencing; COPRO-Seq) of DNA prepared from fecal samples collected over the course of each experiment was used to define the relative abundance of each community member.

When considered as a whole, each mutant library acted similarly to the parental wild-type strain in terms of its proportional abundance in the community (Fig. S4B, Table S3). In both monotonously fed treatment groups, the relative abundance of *B. cellulosilyticus* WH2 was high, although significantly greater in animals fed LF/HPP chow (36±1.6% versus 16±3.1% after 16 days on the diets; p<0.01, Mann-Whitney test). *B. ovatus* ATCC 8483 had a striking preference for conditions associated with LF/HPP chow feeding (11±0.7% versus 0.8±0.3%; p<0.01, Mann-Whitney test). Strain-specific differences were also evident; *B. thetaiotaomicron* VPI-5482 was more successful in the LF/HPP diet context than *B. thetaiotaomicron* 7330 (9±0.5% versus 0.02±0.04 %; p<0.05, Mann-Whitney test) while the 7330 strain had significantly higher abundance than the VPI-5482 strain when animals were consuming the HF/HS diet (6±1.4% versus 3±0.5%; p<0.05, Mann-Whitney test) (Fig. S5A).

Multi-taxon INSeq allowed us to obtain a genome-wide, gene-level view of the effects of the selective pressures exerted by diets on fitness. Analysis of the relative abundances of all *B. cellulosilyticus* mutants in fecal samples collected at 4, 10 and 16 days post-gavage in the two monotonous diet treatment groups disclosed that by day 4 the mutant population had already manifested diet-specific configurations (PERMANOVA of Bray-Curtis dissimilarities, R^2=0.676, p=0.001). The relative abundances of *B. cellulosilyticus* mutants were remarkably consistent between individually caged mice monotonously fed a given diet, including groups of mice housed in different gnotobiotic isolators (Student's unpaired t-test; fig. S6A). This within-treatment group consistency was sustained as the configuration of the mutant population evolved during the ensuing 11 days of monotonous diet consumption [Fig. 1B, fig. S6B, C; since we compared an INSeq library of a given single species/strain

after *in vivo* selection under different dietary treatments, we used the non-phylogenetic Hellinger metric (11) to calculate the dissimilarity between the libraries within or between treatment groups].

Multi-taxon INSeq of fecal samples collected at the end of 16 days of monotonous diet consumption yielded 550 HF/HS diet-specific *B. cellulosilyticus* fitness determinants and 34 LF/HPP diet-specific determinants: 244 of the 550 HF/HS diet-specific genes had KEGG Orthology (KO) assignments; among this group, there was a significant enrichment of genes belonging to the KEGG categories 'Membrane Transport' (e.g., one operon involved in iron transport, another involved in phosphate transport), 'Metabolism of Cofactors and Vitamins' [e.g., seven genes in the cobalamin biosynthesis pathway, consistent with the view that the capacity to synthesize and utilize cobalamin and other substituted corrins is an important determinant of survival in the gut (9, 12)], and 'Protein folding, sorting and degradation' (see Table S4A for the functional annotations and fitness indices of these genes).

Resilience, the ability of a system to "absorb changes of state variables and parameters, and still persist", has been differentiated in community ecology from stability, "the ability of a system to return to an equilibrium state after a temporary disturbance"(13). The two groups of mice subjected to a reciprocal set of diet oscillations [LF/HPP (4d)→HF/HS (6d)→LF/HPP(6d) or HF/HS(4d)→LF/HPP(6d)→HF/HS(6d)] exhibited consistent and marked diet-specific changes in overall community structure as defined by COPRO-Seq (Fig. 1C,D; also see the patterns of change in relative abundances of *B. cellulosilyticus* and *B. ovatus* in fig. S5A). The 15-member community showed elements of both resilience and stability in the face of brief dietary disturbances. Oscillating the mice between the LF/HPP and HF/HS diets produced substantial changes in the relative proportions of community members, but all persisted through the disturbance (resilience), and perturbed and unperturbed communities converged on very similar states determined by the final diet consumed by the mice. At the end of the experiment, the diets explained 82.6% of the variance between fecal communities (Hellinger distances, PERMANOVA, p=0.001, fig. S7A). The interaction between diets and treatment (monotonous diet versus diet oscillation) explained only 8.7% of the variance, which is similar to its explanatory power before the disturbance (R^2=8.7%).

Followup INSeq analysis of the same fecal DNA samples used for COPRO-Seq afforded an opportunity to simultaneously characterize the degree to which the aggregate collections of *B. cellulosilyticus* and *B. ovatus* mutants were able to persist and recover from a diet disturbance (an INSeq-based definition of resilience and stability) and the variation in such recovery between animals (an INSeq-based measure of stochasticity). The *B. cellulosilyticus* WH2 INSeq library showed high resilience to the brief dietary perturbations in the sense that the library as a whole persisted throughout the experiment, though some individual mutants did not. However, it did not show the same stability that the 15-member community did; in the mice that experienced diet oscillations, Hellinger distances indicate that the library of mutants did not converge to the same state as in monotonously fed mice but rather persisted as a reconfigured population. This can be seen in the Hellinger distances between the libraries at the end of the experiment, when diet explained 38.7% of the variance, but the interaction between diet and treatment explained 35.1% (PERMANOVA, p=0.001, fig. S7B) - a value much greater than its explanatory power before the diet oscillation (R^2=8.2%).

The configuration of the *B. cellulosilyticus* mutant population exhibited diet-specific changes. However, during the 'recovery phase', when these two groups of mice were returned to the starting diet of the sequence, the *B. cellulosilyticus* mutant library shifted towards a state similar but not identical to that observed in monotonously fed hosts (Fig. 1B). Of the 550 *B. cellulosilyticus* genes identified as HF/HS specific fitness determinants in monotonously fed animals, 251 had significantly higher fitness indices (i.e., were less required) in animals that had been switched temporarily to LF/HPP chow and then returned to the HF/HS diet compared to mice that had only consumed the HF/HS diet [includes the 6 genes comprising the only one of its 113 polysaccharide utilization loci (PULs) identified as a HF/HS diet-specific fitness determinant (see Table S4A and below)]. Similarly, 21 of the 34 genes designated as LF/HPP-specific fitness determinants in monotonously fed mice had significantly higher fitness indices in mice that had been temporarily switched to the HF/HS diet and then returned to LF/HPP chow (Table S4A). As with *B. cellulosilyticus*, diet oscillations of the type A to B to A (or B to A to B) led to increased similarity in the overall patterns of fitness in the *B. ovatus* mutant library compared to those that were monotonously fed (Fig. S8; Table S4B). This response to perturbation may represent an important mechanism behind the maintenance of diet-specific traits in a member or members of a microbiota harbored by hosts experiencing more varied diets (here we are using mutants as proxies for functional 'traits').

Identifying core *in vivo* fitness determinants in four *Bacteroides* strains

Genes that are conserved among the four *Bacteroides* strains and that show a significant effect on fitness in all strains in both dietary contexts can be defined as a core set of *in vivo* fitness determinants for these members of this genus; as such they inform us about the resource requirements and selective pressures these taxa experience in the gut in this community context and under these two dietary conditions.

In total, 2,238 genes are conserved among all four *Bacteroides* strains. Multi-taxon INSeq of fecal samples, collected from mice colonized for 16d and monotonously fed one or the other diet revealed 82 conserved genes with significant diet-independent effects on fitness in all four strains (Table S5). The fitness indices (*z*-scores) for these genes varied as a function of strain but less so across the different diets ($p<0.001$; two-way ANOVA). Among the 82 core fitness determinants, 15 were components of biosynthesis pathways for arginine (ArgB, ArgC, ArgD, ArgF', ArgE, ArgG, ArgH), aspartate (AspC1), lysine (LysA), methionine (MetA), aromatic amino acids (AroH-TyrA, TyrA2, TyrB), branched chain amino acids (LeuB) and histidine (HisI) (Fig. S9A, Table S6A and Supplemental Results). In follow-up studies, the *B. cellulosilyticus* mutant library was grown to stationary phase in minimal medium lacking amino acids and containing one of several carbon sources [glucose, xylose, arabinose, or wheat arabinoxylan (the most common hemicellulose in cereals)], or in control rich medium [tryptone-yeast extract-glucose (TYG) ($n=6$ replicates/growth condition)]. INSeq analysis of the input library and stationary phase cultures (Table S7) confirmed that Tn mutagenesis of the 15 core fitness determinants annotated as being involved in the biosynthesis of 11 amino acids precluded growth in any of four types of amino acid-depleted minimal media but did not affect growth in the rich medium (see Supplemental Results for further details). Together, these results emphasize the important contribution of amino acid

biosynthesis to the survival of these *Bacteroides* strains in the niches they occupy in the 15-member community in the two diet contexts tested in this study. This is consistent with our previous observation made in gnotobiotic mice that the biomass of a 10-member community composed of all wild-type human gut-derived strains, including the three *Bacteroides* used in the current study (*B. thetaiotaomicron* VPI-5482, *B. ovatus* ATCC 8483, and *B. caccae* ATCC 43185), correlates with the amount of protein in the diet (8).

Genes related to carbohydrate utilization/metabolism were also prominently represented among the 82 shared *in vivo* fitness factors (Fig. S9B, Table S6B). They include: (i) members of arabinose, fructose, glucose/galactose and hexuronate utilization pathways [three inner membrane monosaccharide transporters - glucose/galactose permease (GlcT), arabinose permease (AraP) and fructose transporter (FruP) - plus a novel arabinose-dependent repressor from the NrtR family (AraR), an enzyme from the arabinose catabolic pathway (AraA), two galactose catabolic enzymes (GalM, GalE) and two enzymes from the hexuronate utilization pathway (KdgK, KdgA)]; (ii) three sugar-responsive transcriptional regulators, including homologs of BT4338, a Crp-like transcription factor that controls a proposed global sugar catabolic regulon (14) encompassing up to 30 genes/genome that participate in the utilization of arabinose, xylose, fucose, galacturonate, pectin and β-hexosamines; and (iii) cytoplasmic enzymes involved in starch and glycogen synthesis [starch/glycogen synthase (BT1294), glycogen branching enzymes (BT0771 and BT4303), and alpha-glucanotransferases (BT4304, BT4305)].

Group B vitamins are essential micronutrients that serve as cofactors (or their precursors) used by various enzymes involved in many facets of cellular metabolism. All three *Bacteroides* species are prototrophs for all B-vitamins and their respective cofactors except cobalamin (vitamin B12) which is not synthesized by *B. thetaiotaomicron* or *B. ovatus* (see Supplemental Results for application of the metabolic subsystem approach implemented in SEED (15) to investigate the ability of the three *Bacteroides* species to produce and salvage thiamine, riboflavin, niacin, biotin, pyridoxine, cobalamin, pantothenate, and folate). All three *Bacteroides* species lack orthologs for all known transporters for B-vitamins other than thiamine (vitamin B1) and cobalamin (Fig. S9C, Table S6C). Our *in vivo* multi-taxon INSeq analysis revealed that components of B-vitamin biosynthetic pathways as well as thiamine and cobalamin transporters functioned as strain-specific but not core fitness determinants (see Supplemental Results). This result indicated that in our gnotobiotic model, all three *Bacteroides* species, represented by four strains, have different requirements for B-vitamins and/or deploy divergent strategies for acquiring them (e.g., through *de novo* biosynthesis or from the diet, or from other microbes with B-vitamin transporters) (see Supplemental Results for detailed analysis of *in vitro* fitness determinants in vitamin B/cofactor biosynthetic pathways).

One approach that has been used to search for genes that act as fitness determinants *in vivo* is to identify those that exhibit significant differences in their expression *in vitro* versus in a given body habitat (16). Therefore, we compared RNA-Seq datasets from the four *Bacteroides* strains, generated using fecal samples collected at the same time as samples for INSeq, with RNA-Seq datasets produced from *B. cellulosilyticus* during log phase growth in rich medium or in minimal medium containing single carbon sources including arabinose,

arabinoxylan, xylose, or glucose. The results revealed that only 3 of the 82 core fitness determinants had significantly higher levels of expression *in vivo* under both diet conditions than under any of the *in vitro* conditions examined (n=4 mice/diet treatment; n=3 biological replicates/*in vitro* growth condition) (Fig. S10A). The fact that expression of the remaining 79 core fitness determinants, including the regulator of the proposed sugar catabolic regulon (Fig. S10B), was not higher *in vivo* than in *in vitro* highlights one of the benefits of INSeq, and provides a cautionary note that sole reliance on 'differential expression' as a criterion may miss genes that are critical for survival in a given environmental context.

Differences in fitness determinants in the two *B. thetaiotaomicron* strains

B. thetaiotaomicron VPI-5482 is one of the most-studied human gut *Bacteroides* (4, 5, 14, 17–19). *B. thetaiotaomicron* 7330 is a strain recovered from a healthy adult. Fig. 2A illustrates syntenic regions in their genomes. Table S8A describes the representation of genes in KEGG categories and KEGG pathways for the two strains. Table S8B provides a list of the 91 PULs encompassing 808 predicted ORFs in the 7330 strain, 18 of which (136 ORFs) are unique (i.e., their protein products had <90% similarity to any protein in the predicted VPI-5482 proteome).

The 7330 strain is more successful in the 15-member community context when mice are consuming the HF/HS diet. While the relative representation of each *B. thetaiotaomicron* strain in the 15-member community varied by 340-fold depending upon the diet, their total relative abundance in the community was the same across the two diets (Fig. 2B), raising the possibility that the strains might compete for a niche in this community and that occupancy of this niche by one strain or the other is highly sensitive to diet. Therefore, we used INSeq to identify factors that define their distinct dietary niches (Table S9; includes the 385 HF/HS diet-specific fitness determinants in *B. thetaiotaomicron* 7330). In the absence of dietary polysaccharides, *B. thetaiotaomicron* VPI-5482 adaptively forages on host mucosal glycans suggesting that the capacity to consume both dietary and host-derived glycans is one of the mechanisms by which it is able to survive in the gut (5, 20, 21). The HF/HS diet-specific and 7330 strain-specific fitness factors include genes in two operons specifying components of cation efflux systems plus several glycoside hydrolases predicted to be involved in the catabolism of mucosal glycans (Fig. 2C,D; Table S9B). The latter include genes (*Btheta7330_01433*, *Btheta7330_02903* and *Btheta7330_02906*) that encode three members of glycoside hydrolase family GH20 (β-hexosaminidase, EC 3.2.1.52) that likely cleave GlcNAc residues. Their orthologs are also represented in the VPI-5482 genome but in this strain they do not convey a significant fitness effect. Thus, one functional distinction between the two strains revealed by multi-taxon INSeq is that they likely use different strategies for adaptive foraging of mucosal glycans in the absence of dietary polysaccharides.

Individual *Bacteroides* species contain multiple capsular polysaccharide biosynthesis (*CPS*) loci, allowing for a large number of combinations of expressed loci in different environmental contexts. We used multi-taxon INSeq to identify *CPS* loci and their component genes that were critical for fitness *in vivo* within and across species and strains as a function of diet. The two *B. thetaiotaomicron* strains vividly illustrate the unique

contributions of individual *CPS* loci to survival. *CPS4* was the only one of the VPI-5482 strain's seven *CPS* loci that functioned as a significant fitness determinant in the context of the 15-member community and either of the two diets (as shown in fig. S11B, 20 of its 21 genes, collectively covered by >1200 Tn mutants, were defined as essential). In contrast *CPS3*, a locus in the 7330 strain's genome composed of 18 genes not found in the VPI-5482 strain, was the only one among its six *CPS* loci critical for fitness in either diet context (all 18 genes, covered by >850 mutants, had significant fitness indices; Fig. S11C). The importance of these loci is not simply attributable to their levels of expression: microbial RNA-Seq of fecal samples collected at 4dpg and 16dpg revealed that neither *CPS4* in *B. thetaiotaomicron* VPI-5482 nor *CPS3* in *B. thetaiotaomicron* 7330 is the dominantly expressed locus in either diet context (Table S10). Finally, INSeq established that this strain-specificity for *CPS* fitness effects was evident along the length of the gut (data not shown).

Identifying a diet supplement that can specifically manipulate *B. cellulosilyticus* abundance

B. cellulosilyticus WH2 is equipped with more carbohydrate active enzymes (CAZYmes) dedicated to glycan digestion than any other sequenced Bacteroidetes genome reported to date (7). Its 510 CAZYmes, comprising glycoside hydrolases, polysaccharide lyases, carbohydrate esterases, and associated, non-catalytic carbohydrate binding modules, are distributed among 113 PULs. Remarkably, only one of the 113 PULs (*BcellWH2*_04321-4327) functioned as a significant fitness determinant in the HF/HS diet context (Fig. 3A). The fitness indices of the 127 Tn mutants that mapped to the six genes in this PUL were remarkably consistent between individually caged animals as a function of their diet. Moreover, as noted above, the relative abundances of these mutants were higher in mice that had undergone an HF/HS-LF/HPP-HF/HS diet oscillation than in mice monotonously fed the HF/HS diet (Fig. 3A). This PUL is contained within a region of the genome spanning *BcellWH2_04292-BcellWH2_04327* that contains three hybrid two-component systems (see Supplemental Results and fig. S12 for an analysis of putative HTCS regulons in this genomic region).

The *BcellWH2_04321-4327* PUL consists of a hybrid two-component system (HTCS) transcriptional regulator (DNA binding motifs and regulated operons analyzed in Supplemental Results) and a xylan utilization system core (22) consisting of a SusC/D pair (BcellWH2_04325/26), a hypothetical protein (BcellWH2_04324), a predicted xylanase belonging to glycoside hydrolase family 10 (GH10; BcellWH2_04323), a predicted β-galactosidase from GH35 (BcellWH2_04322), and a predicted feruloylesterase (BcellWH2_04321) (Fig. 3A, fig. S12A,B). GC/MS analysis of the products of acid hydrolysis of the HF/HS diet revealed small amounts of xylose and arabinose (61.2±7.5 μg/g and 62.8±6.2 μg/g, respectively). Our previous *in vitro* growth studies of *B. cellulosilyticus* WH2 cultured in minimal medium supplemented with one of 31 distinct carbohydrate substrates, plus RNA-Seq analysis of the bacterium recovered at mid-log phase from those minimal media that supported its growth, revealed that this PUL was induced by xylan and arabinoxylan (7; Fig. S13A). Arabinoxylans are made of a xylan backbone with α-L-arabinose sidechains. They also contain ferulic acid and other phenolic acids that are

covalently linked to them via ester linkages. A likely role for the product of BcellWH2_04321 family member in the context of this specific PUL would be removal of the ferulic esters from arabinoxylan. The GH10 xylanase targets the xylan backbone, while the GH35 enzyme likely removes the α-linked L-arabinose side chains (α-L-arabinose is identical to β-D-galactose except for the C-6 moiety).

B. ovatus, the only other *Bacteroides* strain in the artificial community that could grow in minimal medium containing purified arabinoxylan as the sole carbon source (Fig. S13B) contains a PUL (*Boavatus_01723-32*) that is induced when this medium is supplemented with arabinoxylan (or xylan) (6). Multi-taxon INSeq of fecal samples collected from mice harboring the 15-member community showed that unlike the xylan and arabinoxylan-inducible *B. cellulosilyticus* PUL *BcellWH2_04321-27*, this PUL is not required for survival of *B. ovatus* in the HF/HS diet context (i.e., Tn mutants in its SusC and xylanase genes produced no significant fitness defects; fig. S13C).

Given these results, we examined the effects of consumption of arabinoxylan purified from wheat on the relative abundances of *B. cellulosilyticus* and *B. ovatus* and their fitness determinants. Our rationale was that we could use arabinoxylan to induce expression of the one *B. cellulosilyticus* WH2 PUL that was a key fitness factor in the HF/HS diet context, and that PUL induction would be accompanied by improved fitness. Therefore, prior to colonization with the 15-member community containing the INSeq libraries, one group of germ-free mice received the HF/HS diet plus drinking water supplemented with wheat-derived arabinoxylan (7.5% w/v) for seven days. Another group received the same diet but without arabinoxylan in their drinking water. All groups were then gavaged with the artificial community: the group pre-treated with arabinoxylan continued to receive the HF/HS diet and supplemented water *ad libitum* for the next 14 days and then was switched to unsupplemented water. A reciprocal treatment group received unsupplemented water plus the HF/HS diet for 14 days followed by a switch to arabinoxylan-supplemented water. A third group received unsupplemented water plus the low-fat, high-plant polysaccharide (LF/HPP) diet for 14 days followed by a switch to supplemented water (while being maintained on the LF/HPP chow); these animals served as a control since the targeted PUL only manifests itself as a fitness determinant in the context of the high-fat, polysaccharide-deficient diet (n=5 individually caged mice per treatment group; fig. 3B).

COPRO-Seq analysis of fecal samples collected at the end of each 14 day treatment period demonstrated that arabinoxylan produced a significant increase in *B. cellulosilyticus* abundance in mice fed the HF/HS diet (adjusted p-value <0.001; Student's t-test; comparing group B to group A at the 14 day time point, and the day 30 versus day 14 time points within group A in fig. 3C), but no significant effect in the LF/HPP diet context (see the day 30 versus day 14 time points for group C). Consistent with the INSeq results showing that the arabinoxylan utilization PUL in *B. ovatus* is not a fitness determinant in the HF/HS diet context, we observed no significant effects of arabinoxylan treatment on the relative abundance of this community member (Fig. 3C). We confirmed these findings in a separate experiment where two groups of gnotobiotic mice harboring the 15-member community were treated for 56 days with a HF/HS diet with or without supplementation of their drinking water with 15% (w/v) arabinoxylan. In this higher dose, longer duration,

monotonous diet experiment, the abundance of *B. cellulosilyticus* WH2 increased significantly (as did levels of cecal short chain fatty acids and deconjugated bile acids), while *B. ovatus* showed no response (Fig. S14A,B). Arabinoxylan treatment did not produce a statistically significant difference in total body weight (Student's *t*-test).

We identified 407 *B. cellulosilyticus* WH2 genes that functioned as fitness determinants when mice consumed the HF/HS diet but not when mice also received arabinoxylan in their drinking water (Table S11A). The arabinoxylan- and xylan-inducible PUL, *BcellWH2_04321-4327* was included among this group of genes (Fig. S15A). *B. cellulosilyticus* contains another PUL encoding a predicted xylanase (*BcellWH2_04296-4307*) that is also induced by xylan and arabinoxylan *in vitro*. Neither of these two PULs are fitness determinants when *B. cellulosilyticus* is grown in minimal medium containing arabinoxylan as the sole carbon source (Table S7). The *BcellWH2_04296-4307* PUL is not a fitness determinant when mice are consuming the HF/HS diet, with or without arabinoxylan, suggesting that together these two PULs - *BcellWH2_04321-04327* (the HF/HS diet specific fitness PUL that is no longer required with arabinoxylan supplementation) and *BcellWH2_04296-4307* (not required under either of the two conditions) - provide redundant yet distinguishable functions in the 15-member community context. A recent study of the orthologous PULs in *B. ovatus* ATCC 8483 showed that they are regulated by different xylan structures: all xylans contain a conserved β1,4-xylose core backbone that is variably decorated with monosaccharide side chains; *Bovatus_01727-01732*, the ortholog of *BcellWH2_04321-04327*, is regulated by simpler xylans while *B. ovatus_03715-03733* is regulated by more complex (decorated) xylans (23). This finding together with our results suggest that (i) the xylan present in the HF/HS diet is a simpler type that is processed by the *BcellWH2_04321-4327* PUL, and (ii) adding a xylan of intermediate complexity (wheat arabinoxylan) induces both of the PULs. We postulate that this dual induction creates a functional redundancy with respect to arabinoxylan utilization in *B. cellulosilyticus* WH2 that accounts for our finding that neither PUL is required for fitness in the setting of arabinoxylan supplementation of the HF/HS diet. Targeted mutation and complementation of genes in one, the other, or both *BcellWH2_04321-04327* and *BcellWH2_04296-4307* would provide an opportunity to independently validate the roles of these two PULs in different diet contexts. Once genetic tools have been developed to stably express the one or two proteins required to complement the corresponding single and double knockout strains *in vivo*, these experiments can and should be performed.

There were several transporters that were no longer significant fitness determinants in arabinoxylan-supplemented mice: they include four MFS-type transporters [a nucleoside H+ symporter, plus transporters for galactose/glucose (GlcT), fructose (FruP) and di/tri-peptides], iron and zinc transporters, and two ABC-type antimicrobial peptide transporters (see Fig. S15B for a ranking based on their *q*-scores). Two MFS-type multidrug transporters and one ABC-type antimicrobial peptide transporter were the only *B. ovatus* genes in the KEGG category 'Membrane transport' that lost their significant fitness index scores with arabinoxylan treatment (these transporters are not homologous to those identified in *B. cellulosilyticus*). A number of human gut *Bacteroides*, including *B. ovatus*, are known to be highly resistant to antimicrobial peptides (24). Import of gut mucosal-derived antimicrobial

peptides is one strategy gut symbionts use for avoiding destruction of their cell membranes (25). Some *Bacteroides* species are able to turn to gut mucosal glycans as alterative sources of nutrients when polysaccharides are not well represented in the diet. Arabinoxylan supplementation may mitigate this need and thus the observed diminished requirement for anti-microbial peptide transporters could reflect alterations in interactions with the mucosa.

A total of 27 genes were identified in the *B. cellulosilyticus* genome that functioned as significant fitness determinants when mice received the HF/HS diet plus arabinoxylan but not when animals received the HF/HS diet alone (Table S11A). They include: (i) all genes in an operon (*BcellWH2_00893-895*) that includes a sulfotransferase and sulfate transferase (EC 3.7.74), consistent with the observed increase in deconjugated bile acids documented by UPLC-MS of cecal contents (Fig. S14C), and (ii) several genes that highlight how increases in arabinoxylan availability enhance the importance of ammonium utilization for synthesis of amino acids and proteins in this community and diet context [an ammonium transporter (*BcellWH2_05255*), plus glutamine, and glutamate synthetases (*BcellWH2_5244* and *BcellWH2_05271*, respectively)].

Even though *B. ovatus* did not manifest a significant change in its relative abundance when mice received arabinoxylan supplemented water, multi-taxon INSeq revealed 41 genes that were not fitness determinants when mice consumed the HF/HS diet alone but 'acquired' significant *z*-scores when arabinoxylan was introduced (for a list, see Table S11B). They include seven closely linked glycosyltransferase genes (*Bovatus_03504-11*), suggesting that this organism changes its glycan utilization strategies when it encounters arabinoxylan in the gut environment.

Prospectus

We cannot distinguish a direct effect of a given diet or diet oscillation on a given community member from a primary effect of that diet or diet perturbation on one or more other community members that interact with the member/strain exhibiting changes in its abundance. We could have limited ourselves to mono-colonizations to establish direct effects of diet on these features, but we would have lost our ability to describe these responses and ascertain the niches of these organisms in the more 'natural' context of a microbial community.

In our study, multi-taxon INSeq is used to provide an 'operational' description of an organism's niche by determining which sets of genes allow a bacterial strain to co-exist with other strains/species under a defined set of habitat conditions (e.g., gnotobiotic mice representing a given genetic background, colonized with a given set of sequenced organisms and fed a given set of diets). In principle, this approach can be applied to gut and non-gut habitats in gnotobiotic mice representing different genetic backgrounds, harboring different microbial consortia and manipulated in various ways (diet, immune system, etc.) to obtain a more comprehensive picture of how host genotype influences the niches available. It also offers a way to address a range of questions from how to functionally discriminate strains when exploring the genetic foundations of opportunism, stochasticity in community assembly or the stability and resilience after perturbations, to the selection of probiotic

candidates for bioremediation of perturbed/dysbiotic microbial communities, identifying compounds that affect the functional configurations of a targeted microbiota, and characterization of the nutritional effects/value of diets and their ingredients.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank David O'Donnell and Maria Karlsson for their assistance with gnotobiotic mouse husbandry, Xuefeng Zhou and Ashton Bell for their work on PacBio sequencing, plus Alejandro Reyes, Mark Charbonneau, Philip Ahern, and Ansel Hsiao for many helpful suggestions.

This work was supported in part by grants from the NIH (DK30292, DK70977, DK52574, and GM108527), the Crohn's and Colitis Foundation of America, the European Union's Seventh Framework Program (FP/2007/2013)/ERC Grant Agreement no. 322820 (to B.K.), the Monsanto Company (for PacBio and Illumina sequencing and finishing the *B. thetaiotaomicron* 7330, *B. ovatus* ATCC8483 and *B. cellulosilyticus* WH2 genomes), and the Russian Science Foundation (to D.R. and M.K.)

A patent application (PCT/US2014/045141) has been filed by Washington University related to methods for identifying dietary supplements, such as arabinoxylans, that can be used to deliberately manipulate the representation of targeted members of the human gut microbiota.

References and Notes

1. Human Microbiome Jumpstart Reference Strains. A catalog of reference genomes from the human microbiome. Science. 2010; 328:994–999. [PubMed: 20489017]
2. Qin J, et al. A human gut microbial gene catalogue established by metagenomic sequencing. Nature. 2010; 464:59–65. [PubMed: 20203603]
3. Faith JJ, et al. The long-term stability of the human gut microbiota. Science. 2013; 341
4. Xu J, et al. A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. Science. 2003; 299:2074–2076. [PubMed: 12663928]
5. Sonnenburg JL, et al. Glycan foraging in vivo by an intestine-adapted bacterial symbiont. Science. 2005; 307:1955–1959. [PubMed: 15790854]
6. Martens EC, et al. Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. PLoS Biology. 2011; 9:e1001221. [PubMed: 22205877]
7. McNulty NP, et al. Effects of diet on resource utilization by a model human gut microbiota containing Bacteroides cellulosilyticus WH2, a symbiont with an extensive glycobiome. PLoS Biology. 2013; 11:e1001637. [PubMed: 23976882]
8. Faith JJ, McNulty NP, Rey FE, Gordon JI. Predicting a human gut microbiota's response to diet in gnotobiotic mice. Science. 2011; 333:101–104. [PubMed: 21596954]
9. Goodman AL, et al. Identifying genetic determinants needed to establish a human gut symbiont in its habitat. Cell Host & Microbe. 2009; 6:279–289. [PubMed: 19748469]
10. Goodman AL, Wu M, Gordon JI. Identifying microbial fitness determinants by insertion sequencing using genome-wide transposon mutant libraries. Nature Protocols. 2011; 6:1969–1980. [PubMed: 22094732]
11. Rao, CR. Multivariate Statistics and Matrices in Statistics. Tiit, TKEM.; Niemi, H., editors. Vol. 3. 1995.
12. Degnan PH, Barry NA, Mok KC, Taga ME, Goodman AL. Human gut microbes use multiple transporters to distinguish vitamin B(1)(2) analogs and compete in the gut. Cell Host & Microbe. 2014; 15:47–57. [PubMed: 24439897]
13. Holling CS. Resilience and Stability of Ecological Systems. Annual Review of Ecology and Systematics. 1973; 4:1–23.

14. Ravcheev DA, Godzik A, Osterman AL, Rodionov DA. Polysaccharides utilization in human gut bacterium Bacteroides thetaiotaomicron: comparative genomics reconstruction of metabolic and regulatory networks. BMC Genomics. 2013; 14:873. [PubMed: 24330590]
15. Overbeek R, et al. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res. 2005; 33:5691–5702. [PubMed: 16214803]
16. Lindenstrauss AG, et al. Transcriptome analysis of Enterococcus faecalis toward its adaption to surviving in the mouse intestinal tract. Arch Microbiol. 2014; 196:423–433. [PubMed: 24700373]
17. Reeves AR, D'Elia JN, Frias J, Salyers AA. A Bacteroides thetaiotaomicron outer membrane protein that is essential for utilization of maltooligosaccharides and starch. J Bacteriol. 1996; 178:823–830. [PubMed: 8550519]
18. Reeves AR, Wang GR, Salyers AA. Characterization of four outer membrane proteins that play a role in utilization of starch by Bacteroides thetaiotaomicron. J Bacteriol. 1997; 179:643–649. [PubMed: 9006015]
19. Shipman JA, Berleman JE, Salyers AA. Characterization of four outer membrane proteins involved in binding starch to the cell surface of Bacteroides thetaiotaomicron. J Bacteriol. 2000; 182:5365–5372. [PubMed: 10986238]
20. Martens EC, Chiang HC, Gordon JI. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. Cell Host & Microbe. 2008; 4:447–457. [PubMed: 18996345]
21. Kashyap PC, et al. Genetically dictated change in host mucus carbohydrate landscape exerts a diet-dependent effect on the gut microbiota. Proc Natl Acad Sci USA. 2013; 110:17059–17064. [PubMed: 24062455]
22. Dodd D, Mackie RI, Cann IK. Xylan degradation, a metabolic property shared by rumen and human colonic Bacteroidetes. Mol Microbiol. 2011; 79:292–304. [PubMed: 21219452]
23. Rogowski A, et al. Glycan complexity dictates microbial resource allocation in the large intestine. Nat Commun. 2015; 6:7481. [PubMed: 26112186]
24. Cullen TW, et al. Gut microbiota. Antimicrobial peptide resistance mediates resilience of prominent gut commensals during inflammation. Science. 2015; 347:170–175. [PubMed: 25574022]
25. Shelton CL, Raffel FK, Beatty WL, Johnson SM, Mason KM. Sap transporter mediated import and subsequent degradation of antimicrobial peptides in Haemophilus. PLoS Pathog. 2011; 7:e1002360. [PubMed: 22072973]
26. Krzywinski M, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009; 19:1639–1645. [PubMed: 19541911]
27. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One. 2010; 5:e11147. [PubMed: 20593022]
28. Chin CS, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods. 2013; 10:563–569. [PubMed: 23644548]
29. Reyes A, Wu M, McNulty NP, Rohwer FL, Gordon JI. Gnotobiotic mouse model of phage-bacterial host dynamics in the human gut. Proc Natl Acad Sci USA. 2013; 110:20236–20241. [PubMed: 24259713]
30. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature Methods. 2012; 9:357–359. [PubMed: 22388286]
31. Ning Z, Cox AJ, Mullikin JC. SSAHA: a fast search method for large DNA databases. Genome Res. 2001; 11:1725–1729. [PubMed: 11591649]
32. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010; 11:R106. [PubMed: 20979621]
33. Rey FE, et al. Metabolic niche of a prominent sulfate-reducing human gut bacterium. Proc Natl Acad Sci USA. 2013; 110:13582–13587. [PubMed: 23898195]
34. Lehr T, Geyer H, Maass K, Doenhoff MJ, Geyer R. Structural characterization of N-glycans from the freshwater snail Biomphalaria glabrata cross-reacting with Schistosoma mansoni glycoconjugates. Glycobiology. 2007; 17:82–103. [PubMed: 16971380]
35. Saier MH Jr, Reddy VS, Tamang DG, Vastermark A. The transporter classification database. Nucleic Acids Res. 2014; 42:D251–D258. [PubMed: 24225317]

36. Hudson AO, et al. Dual diaminopimelate biosynthesis pathways in Bacteroides fragilis and Clostridium thermocellum. Biochim Biophys Acta. 2011; 1814:1162–1168. [PubMed: 21616177]

37. Novichkov PS, et al. RegPrecise 3.0--a resource for genome-scale exploration of transcriptional regulation in bacteria. BMC Genomics. 2013; 14:745. [PubMed: 24175918]

38. Xu J, Chiang HC, Bjursell MK, Gordon JI. Message from a human gut symbiont: sensitivity is a prerequisite for sharing. Trends in Microbiol. 2004; 12:21–28.

39. Martens EC, Roth R, Heuser JE, Gordon JI. Coordinate regulation of glycan degradation and polysaccharide capsule biosynthesis by a prominent human gut symbiont. J Biol Chem. 2009; 284:18445–18457. [PubMed: 19403529]

40. Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS. Comparative genomics of thiamin biosynthesis in procaryotes. New genes and regulatory mechanisms. J Biol Chem. 2002; 277:48949–48959. [PubMed: 12376536]

41. Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS. Comparative genomics of the vitamin B12 metabolism and regulation in prokaryotes. J Biol Chem. 2003; 278:41148–41159. [PubMed: 12869542]

42. Gerdes SY, et al. Experimental determination and system level analysis of essential genes in Escherichia coli MG1655. J Bacteriol. 2003; 185:5673–5684. [PubMed: 13129938]

43. Rodionov DA. Comparative genomic reconstruction of transcriptional regulatory networks in bacteria. Chem Rev. 2007; 107:3467–3497. [PubMed: 17636889]

44. Mironov AA, Vinokurova NP, Gel'fand MS. Software for analyzing bacterial genomes. Mol Biol (Mosk). 2000; 34:253–262. [PubMed: 10779952]

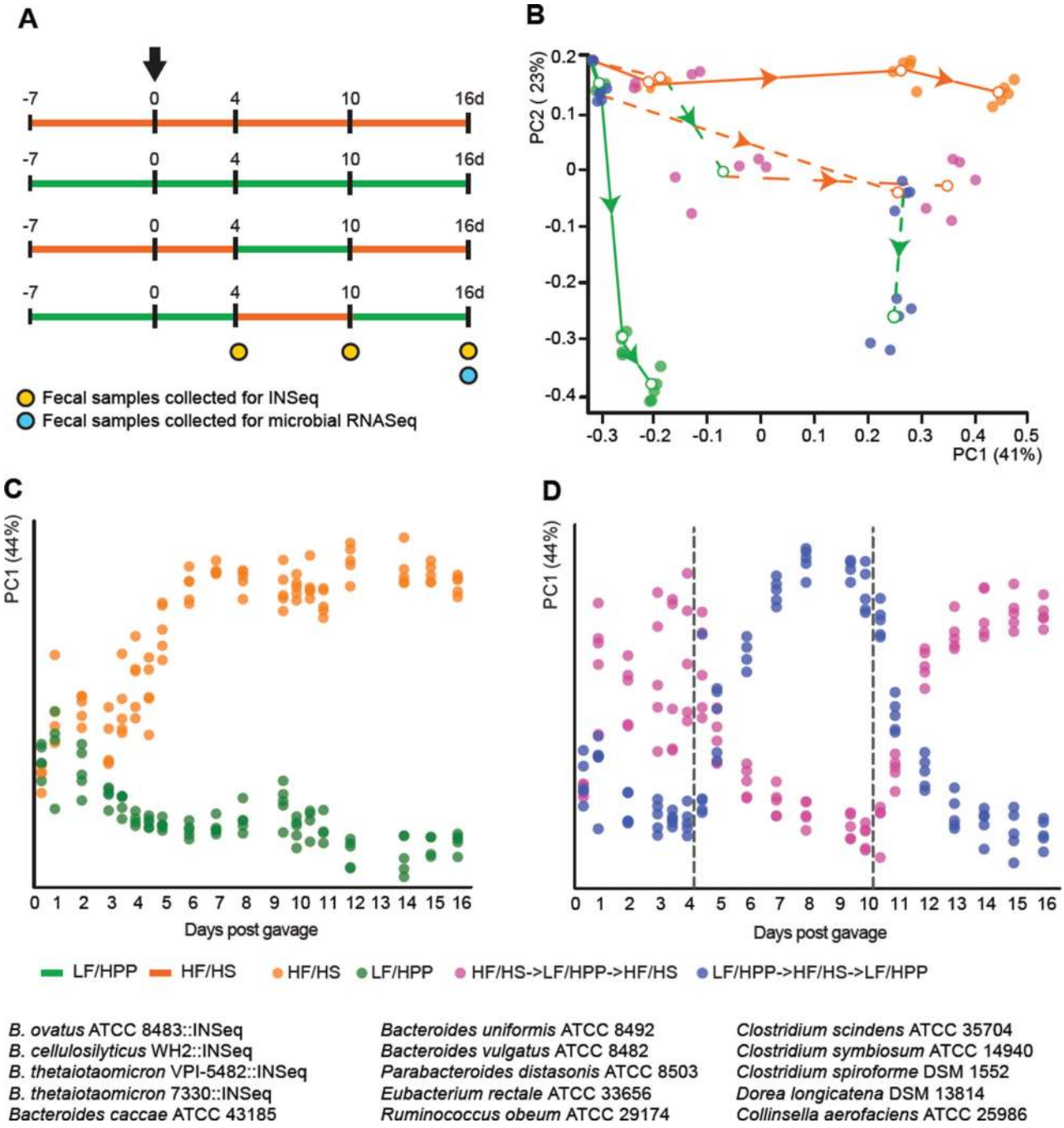

Fig. 1. The effect of diet and diet oscillations on the configuration of the 15 member artificial community containing 11 wild-type strains and the four mutant libraries, including the structure of the population of *B. cellulosilyticus* WH2 mutants

(**A**) Experimental design. 10 to 12-week-old germ-free male C57BL/6J mice were gavaged with the indicated consortium of 15 strains. Animals were given the LF/HPP and/or the HF/HS diet in the order shown. Fecal samples were collected at the indicated time points for INSeq and microbial RNA-Seq analyses. (**B**) Principal coordinates analysis (PCoA) of Hellinger distance measurements based on the relative abundance of *B. cellulosilyticus* WH2 Tn mutants in fecal samples, as defined by multi-taxon INSeq analysis. Each colored circle

represents the population of all mutants in *B. cellulosilyticus* WH2 sampled from an individual mouse in the indicated diet treatment group at a given time point. Each empty circle represents the average location of the mutant population for a given group at 4, 10 and 16 days post gavage. The lines connect time points in a given group. The color of the lines indicates the diet. Dashed lines indicate the diet oscillation groups. (**C,D**) PCoA of Hellinger distance measurements based on COPRO-Seq data. Each colored circle represents a fecal community sampled from an individually caged mouse belonging to the indicated diet treatment group. These circles were ordinated in the same coordinate space but are being shown as two separate panels for clarity. Vertical dashed lines in panel D indicate the time point when a diet switch occurred.

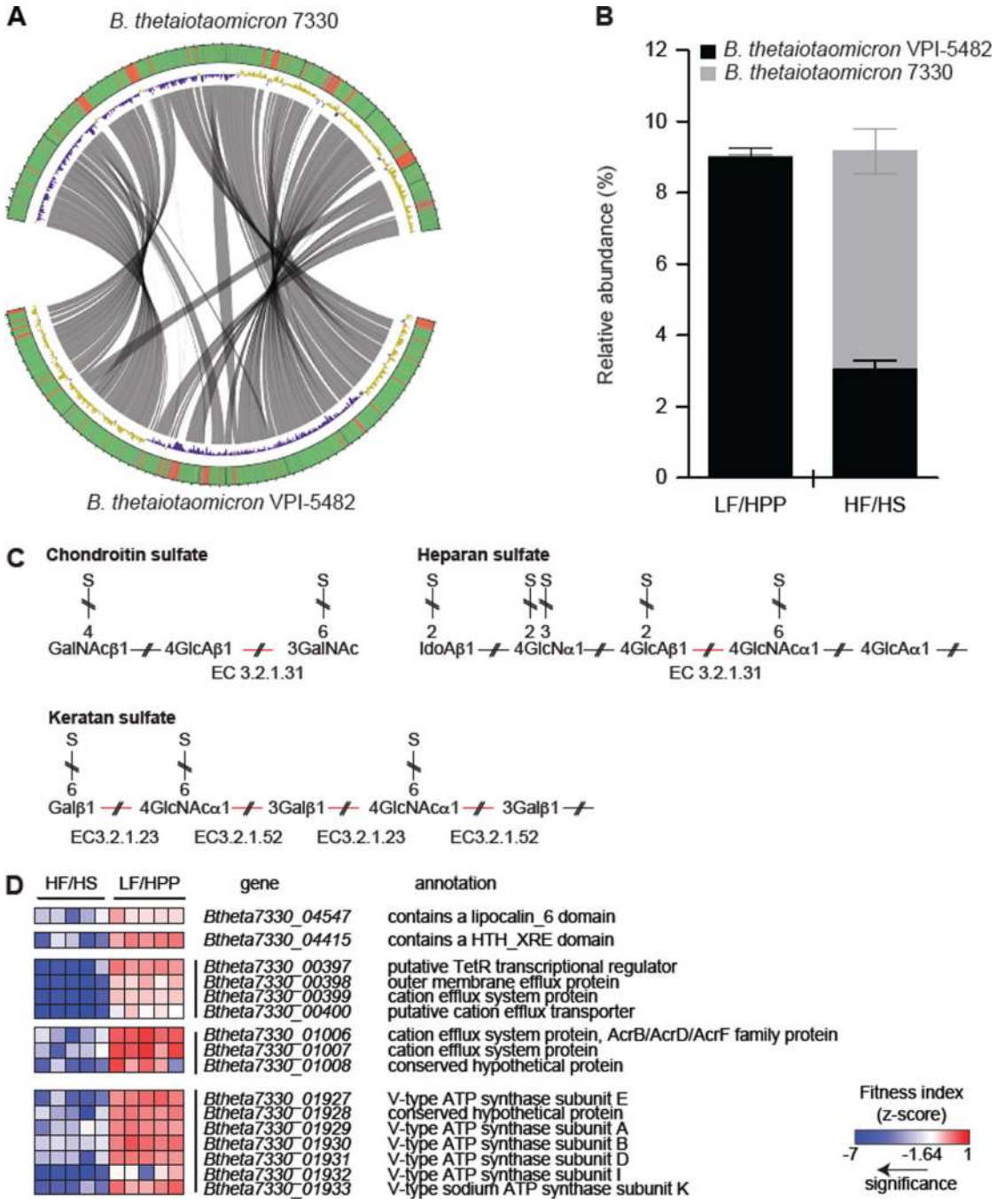

Fig. 2. HF/HS diet-specific fitness determinants in *B. thetaiotaomicron* 7330

(**A**) Circos (26)-generated alignment of the *B. thetaiotaomicron* VPI-5482 and *B. thetaiotaomicron* 7330 genomes. Grey lines connect segments of DNA conserved between these two genomes (these regions were identified by alignment with Mauve; 27). The color-coded outer circle denotes the similarity between their proteins: green, >90% identity (based on Blastp alignment); blue, 70%-90% identity; red, <70% identity; black, intergenic regions. GC skew was calculated using a sliding window size of 10kb (yellow, GCskew>0; purple, GCskew<0). (**B**) COPRO-Seq analysis of the relative abundance of the two *B.*

thetaiotaomicron strains in the fecal microbiota of mice sampled two weeks after gavage while consuming the HF/HS or LF/HPP diets. Mean values ± SEM are shown (n=5 individually caged mice harboring a community consisting of 11 wild-type and the four mutant libraries/treatment group). Note that the summed relative abundance of the two strains remains the same even though the relative representation of the individual strains is significantly different in the two diet contexts (p<0.001, 2-way ANOVA). (**C**) HF/HS diet-specific fitness determinants in *B. thetaiotaomicron* 7330 involved in degradation of glycosaminoglycans associated with the intestinal mucosa (genes highlighted with red lines together with their EC annotations). (**D**) HF/HS diet-associated changes in the z-scores of 7330-strain-specific fitness determinants that are involved in transcription regulation (*Btheta7330_04415*) or are components of operons encoding transport systems (vertical black lines to the left of the gene ID denote individual operons).

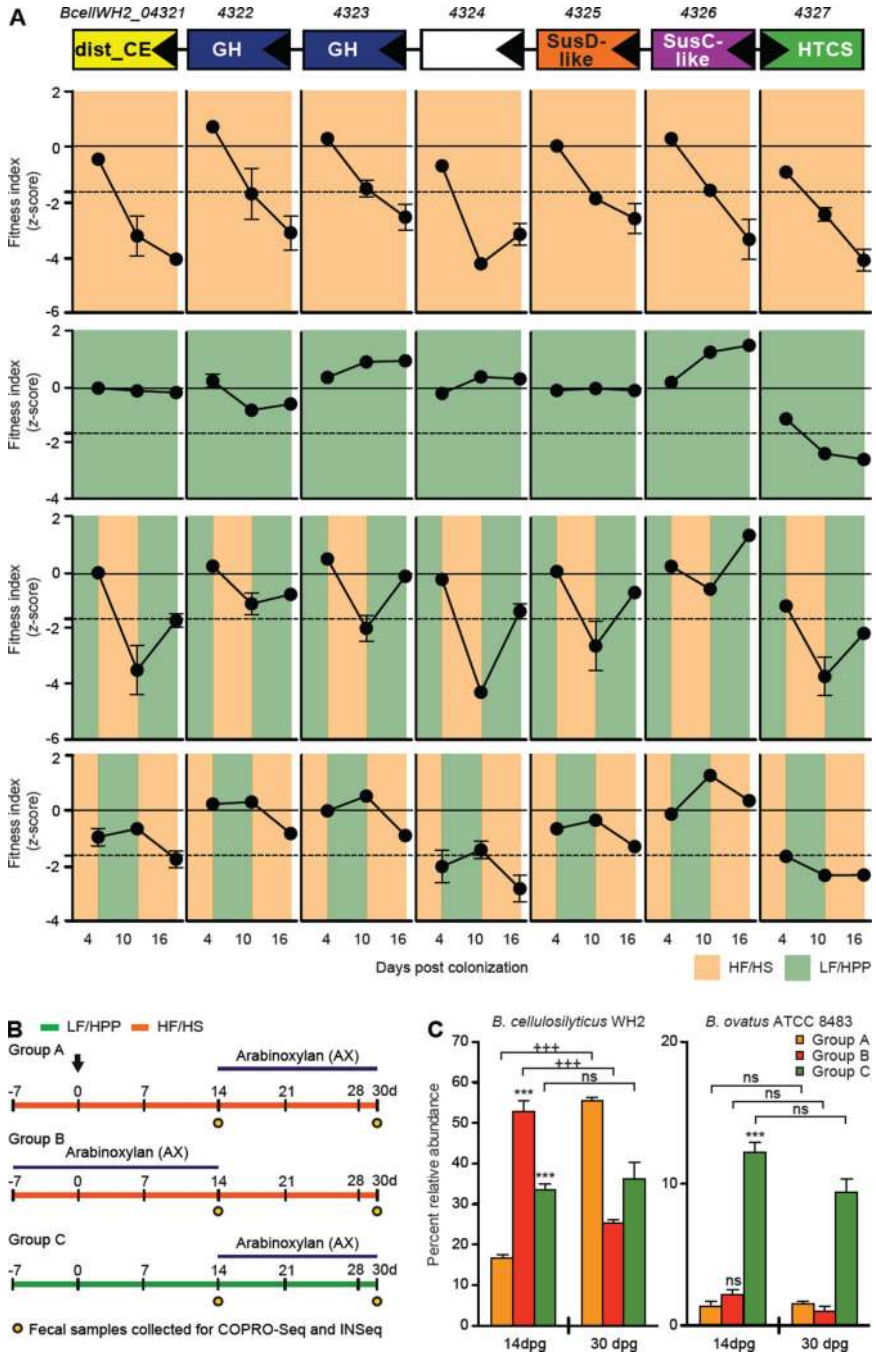

Fig. 3. Arabinoxylan increases the relative abundance of *B. cellulosilyticus* WH2 *in vivo*

(**A**) INSeq analysis reveals that all genes in PUL *BcellWH2_04321-4327* have significant fitness indices (*z*-values) in the HF/HS diet context. *BACWH2_04327*, encoding a hybrid two-component system (HTCS) regulator, is the only gene in this PUL that has a significant fitness effect on the LF/HPP diet. Functional annotations for genes in the PUL are shown together with the direction of their transcription. Fitness indices for each gene in the different diet contexts (orange, HF/HS; green LF/HPP) are plotted as mean values ± SEM. The horizontal dashed line indicates the cutoff for significance ($p<0.05$; *z*-test with FDR

correction). Abbreviations: dist_CE, distant relatives of carbohydrate esterase; GH, Glycoside Hydrolase; HTCS, Hybrid Two-Component System. (**B**) Experimental design. Adult C57BL/6J germ-free mice were gavaged with a consortium containing 11 wild-type strains plus the four *Bacteroides* INSeq libraries. Animals were fed the HF/HS or LF/HPP diets with or without supplementation of their drinking water with 7.5% arabinoxylan (n=5 individually-caged mice/group). (**C**) The relative abundance of *B. cellulosilyticus* and *B. ovatus* was defined by COPRO-Seq analysis of fecal samples collected at the indicated time points. Mean values ± SEM are plotted. ***, p <0.001 compared to the reference group A at 14 days post-gavage (dpg); +++, p <0.001 for within group comparisons of the 30 dpg versus 14 dpg fecal sample (Student's t-test after FDR correction). *B. ovatus*, the only other *Bacteroides* strain in the community that exhibited significantly increased growth in minimal medium supplemented with arabinoxylan (Fig. S13B) did not manifest a significant change in its relative abundance *in vivo* when arabinoxylan was added to the drinking water (ns, not significant).