



HAL
open science

Hadza Prevotella require diet-derived microbiota-accessible carbohydrates to persist in mice

Rebecca Gellman, Matthew Olm, Nicolas Terrapon, Fatima Enam, Steven Higginbottom, Justin Sonnenburg, Erica Sonnenburg

► **To cite this version:**

Rebecca Gellman, Matthew Olm, Nicolas Terrapon, Fatima Enam, Steven Higginbottom, et al.. Hadza Prevotella require diet-derived microbiota-accessible carbohydrates to persist in mice. *Cell Reports*, 2023, 42 (11), pp.113233. 10.1016/j.celrep.2023.113233 . hal-04461077

HAL Id: hal-04461077

<https://cnrs.hal.science/hal-04461077v1>

Submitted on 28 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

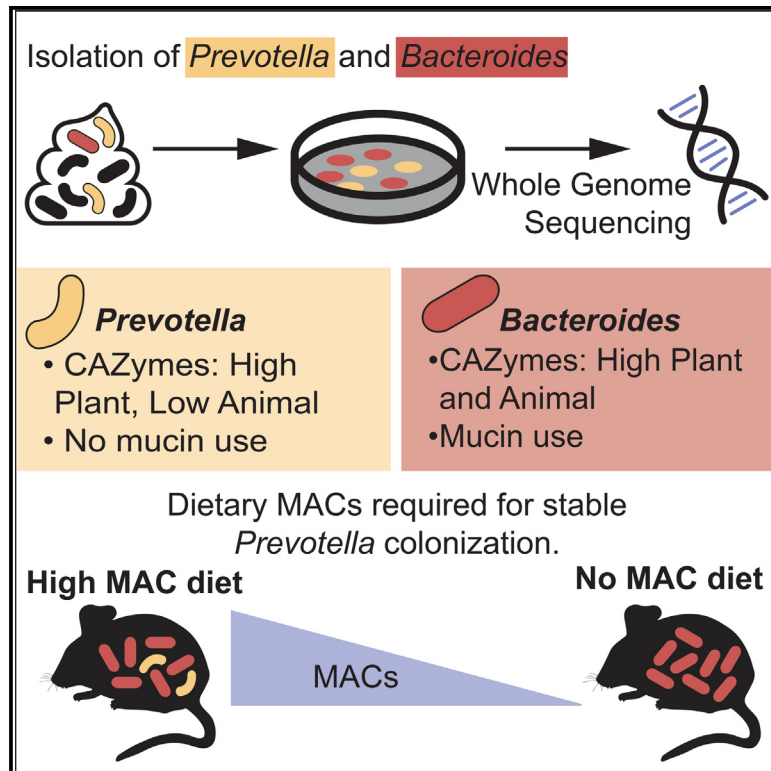
L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Hadza *Prevotella* require diet-derived microbiota-accessible carbohydrates to persist in mice

Graphical abstract



Authors

Rebecca H. Gellman, Matthew R. Olm, Nicolas Terrapon, Fatima Enam, Steven K. Higginbottom, Justin L. Sonnenburg, Erica D. Sonnenburg

Correspondence

jsonnenburg@stanford.edu (J.L.S.), erica.sonnenburg@stanford.edu (E.D.S.)

In brief

Gellman et al. present a set of *Bacteroides* and *Prevotella* isolates from the Hadza microbiota. The results of whole-genome sequencing, gnotobiotic mouse models, and RNA sequencing show that *P. copri* relies on the presence of dietary microbiota-accessible carbohydrates (MACs) to persist in the gut microbiota.

Highlights

- *Bacteroides* and *Prevotella* sp. are isolated, sequenced from Hadza fecal samples
- *Bacteroides* sp. encode more mucus degrading capacity than *Prevotella* sp.
- Unlike *Bacteroides*, *Prevotella* colonization requires dietary plant fiber
- *Bacteroides* outcompetes *Prevotella* *in vivo* on low plant fiber diet



Article

Hadza *Prevotella* require diet-derived microbiota-accessible carbohydrates to persist in mice

Rebecca H. Gellman,¹ Matthew R. Olm,¹ Nicolas Terrapon,² Fatima Enam,¹ Steven K. Higginbottom,¹ Justin L. Sonnenburg,^{1,3,4,*} and Erica D. Sonnenburg^{1,4,5,*}

¹Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA, USA

²Architecture et Fonction des Macromolécules Biologiques, INRAE, CNRS, Aix-Marseille Université, Marseille, France

³Chan Zuckerberg Biohub, San Francisco, CA, USA

⁴Center for Human Microbiome Studies, Stanford University School of Medicine, Stanford, CA, USA

⁵Lead contact

*Correspondence: jsonnenburg@stanford.edu (J.L.S.), erica.sonnenburg@stanford.edu (E.D.S.)

<https://doi.org/10.1016/j.celrep.2023.113233>

SUMMARY

Industrialization has transformed the gut microbiota, reducing the prevalence of *Prevotella* relative to *Bacteroides*. Here, we isolate *Bacteroides* and *Prevotella* strains from the microbiota of Hadza hunter-gatherers in Tanzania, a population with high levels of *Prevotella*. We demonstrate that plant-derived microbiota-accessible carbohydrates (MACs) are required for persistence of *Prevotella copri* but not *Bacteroides thetaiotaomicron in vivo*. Differences in carbohydrate metabolism gene content, expression, and *in vitro* growth reveal that Hadza *Prevotella* strains specialize in degrading plant carbohydrates, while Hadza *Bacteroides* isolates use both plant and host-derived carbohydrates, a difference mirrored in *Bacteroides* from non-Hadza populations. When competing directly, *P. copri* requires plant-derived MACs to maintain colonization in the presence of *B. thetaiotaomicron*, as a no-MAC diet eliminates *P. copri* colonization. *Prevotella*'s reliance on plant-derived MACs and *Bacteroides*' ability to use host mucus carbohydrates could explain the reduced prevalence of *Prevotella* in populations consuming a low-MAC, industrialized diet.

INTRODUCTION

The industrialized lifestyle is defined by the consumption of highly processed foods, high rates of antibiotic administration, cesarean section births, sanitation of the living environment, and reduced contact with animals and soil, all of which can affect the human gut microbiota.¹ Certain taxa are influenced by industrialization; i.e., they are prevalent and abundant in non-industrialized populations and diminished or absent in industrialized populations, or vice versa.^{2–8} The microbiota of 1,000- to 2,000 year-old North American paleofeces is more similar to the modern non-industrialized than industrialized gut.⁹ The industrialized microbiota appears to be a product of both microbial extinction, as once-dominant taxa disappear, and expansion of less-dominant or new taxa.¹⁰

The industrialized diet differs drastically from non-industrialized diets, including a reduced amount of microbiota-accessible carbohydrates (MACs), a major metabolic input for microbes in the distal gastrointestinal tract.^{10–12} Some gut-resident microbes use host mucin, which is heavily glycosylated, as a carbon source, depending on the availability of dietary MACs.^{13–17} Shifts in dietary MACs alter microbial relative abundances and may increase inflammation and susceptibility to intestinal patho-

gens.^{14,18,19} Taxa are lost due to a lack of dietary MACs over generations in a mouse model²⁰ and in humans as they immigrate to the US.⁷

As human populations adopt an industrialized lifestyle, the prevalence of *Prevotella* decreases and that of *Bacteroides* increases.^{2,3,21} These genera are both members of the Bacteroidota phylum, are known to colonize mammalian hosts, and make up a significant fraction of the human gut microbiome.^{22–24} While *Bacteroides* are well studied, *Prevotella* species remain understudied with few tools available for mechanistic investigation.^{25–28} Both genera harbor well-documented carbohydrate utilization capabilities, encoded in carbohydrate active enzymes (CAZymes), often organized into polysaccharide utilization loci (PULs).^{29–31} Characterization of intestinal *Prevotella* species have been limited by challenges with colonization, particularly mono-colonization of germ-free mice. Here, we overcome these barriers to establish a causal link between diet and *Prevotella copri* abundance in a gnotobiotic mouse model.

The decreased prevalence of *Prevotella* in industrial populations is likely linked to a decline in relative abundance within individual microbiotas.³² Decreased abundance of bacterial taxa in individuals reduces the likelihood of transmission from mother



to infant.^{1,5} When compounded over generations, decreased abundance can result in population-level decline in prevalence and eventually taxa loss or extinction.⁷ The factors driving the decline in *Prevotella* and the increase in *Bacteroides* during industrialization remain elusive. The abundance and prevalence of specific strains of *P. copri*, the dominant *Prevotella* species in the human gut, vary among populations based on host lifestyle, particularly diet.^{33,34} Here, we use gnotobiotic mice to investigate the role of diet in sustaining *Prevotella* and *Bacteroides* colonization; we demonstrate that dietary MACs play a key role in controlling the abundances of *Bacteroides* and *Prevotella*.

RESULTS

Bacteroides and *Prevotella* genomes from the Hadza microbiota vary in prevalence across lifestyle

To compare *Prevotella* and *Bacteroides* from non-industrialized lifestyle populations, we isolated and sequenced six *Bacteroides* strains and seven *Prevotella* strains from stool samples collected from 13 Hadza individuals. Single-isolate genomes were assembled using both MiSeq-generated short reads (146 bp) and nanopore-generated long reads (10–100 kb) (Table 1).

The taxonomy of these newly isolated strains was evaluated using the Genome Taxonomy Database (GTDB) version r207 (Figure 1A). All *Bacteroides* isolates belong to known species: *Bacteroides ovatus*, *Bacteroides thetaiotaomicron*, *Bacteroides caccae*, and *Bacteroides fragilis*. Three of our *Prevotella* isolates belong to named species *Prevotella sp015074785* and *Prevotella sp900551275*, while the remaining five isolates are novel species according to GTDB. To verify this finding, we created a phylogenetic tree with our *Prevotella* isolates, representatives of the most closely related representative species in GTDB, and all *P. copri* representative genomes in GTDB (Figure S1A; Table S1). We observe that our isolated genomes have approximately the same phylogenetic distance to the closest representative genomes as the representative genomes have to one another, supporting their characterization as novel species. Apart from GTDB, there have been other efforts to characterize the extensive genomic diversity of the *Prevotella* genus.³⁴ Of the four proposed *P. copri* subgroups possessing >10% inter-clade genetic divergence, all eight Hadza *Prevotella* strains recovered in this study belong to clade A (Figure S1B).

To understand the prevalence of these genomes across human populations, we compared *Prevotella* and *Bacteroides* prevalence among Hadza adults and infants, four populations from Nepal living on a lifestyle gradient including foraging (Chepang), recent agriculturalist (Raute, Raji), longer term agriculturalist (Tharu), and industrial lifestyle populations (California) (Figure 1B; Table S2). We chose these groups due to their varied lifestyles and the exceptional metagenomic sequencing depth achieved, averaging 23 Gbp per sample.^{3,4} *Prevotella* genomes are rare in or absent from the industrialized populations, while they are more prevalent and abundant in the Hadza and Nepali samples. Conversely, nearly all *Bacteroides* genomes, including those isolated from the Hadza, are more prevalent in the California samples. The clear lifestyle shift associated with *Bacteroides* and *Prevotella* prevalence leads to the question of what aspects of the industrial lifestyle have driven these changes.

Dietary MACs are necessary for *P. copri* persistence

While many factors differentiate the industrial and non-industrial lifestyles, diet serves as the top candidate for driving microbiota alterations.¹⁰ The Hadza diet is rich in dietary MACs from foraged tubers, berries, and baobab.³⁵ In contrast, the industrialized diet is typified by high caloric intake and foods rich in fat and low in MACs.³⁶ We wondered whether diet alone could affect the ability of Hadza *Bacteroides* and *Prevotella* to colonize mice. Germ-free (GF) mice were colonized with either Hadza *B. thetaiotaomicron* (Bt) H-2622, or Hadza *P. copri* (Pc) H-2477. Mice were maintained on a high-MAC diet for 7 days and then switched to either a diet devoid of MACs (no MAC), a high-fat/low-MAC diet (Western), or maintained on the high-MAC diet for 7 days (Figure 2A). Bt H-2622 colonization density (10⁹ colony-forming units [CFU]/mL in feces) at baseline on the high-MAC diet was maintained in all three diet conditions (Figure 2B). Pc H-2477 colonized to a lower degree on the high-MAC diet (10⁷ CFU/mL on day 0) and declined drastically following the change to the Western or no-MAC diet, with no fecal CFUs detectable 7 days post diet switch (Figure 2C). The lack of detectable Pc H-2477 in the absence of MACs was particularly striking given the absence of competition from other microbes in this mono-associated state. To our knowledge, this is the first example of a strain's apparent eradication in a mono-associated state due to a diet change. Two other *P. copri* strains (Hadza Pc H-2497 and a non-Hadza strain isolated from an individual of African origin Pc N-01) are also lost *in vivo* in the absence of dietary MACs (Figures S2A and S2B), indicating that survival of *P. copri* *in vivo* depends on the presence of dietary MACs.

To measure the gene expression employed by Hadza Pc and Bt *in vivo*, we analyzed transcriptional profiling data from cecal contents of mice monocolonized with either Pc H-2477 or Bt H-2622 fed a high-MAC diet relative to *in vitro* growth in peptone yeast glucose broth (PYG). Both Bt H-2622 and Pc H-2477 upregulate a large number of genes *in vivo* under high-MAC diet conditions. Despite the fact that 18% and 13% of genes in the Bt H-2622 and Pc H-2477 genomes, respectively, encode for predicted carbohydrate utilization proteins, 86% (in Bt H-2622) and 65% (in Pc H-2477) of genes upregulated *in vivo* relative to *in vitro* encode for carbohydrate utilization ($p < 4e-12$ for Bt, $p < 5e-13$ for Pc, Fisher's exact test), indicating that carbohydrate utilization is the major metabolic function of these organisms *in vivo* (Figure 2D).

A comparison of glycosidic linkage-breaking CAZymes, glycoside hydrolases (GHs) and polysaccharide lyases (PLs), reveals that both Bt H-2622 and Pc H-2477 upregulate more CAZymes *in vivo* on the high-MAC diet than *in vitro* (Pc, 71/6 CAZymes significantly expressed *in vivo/in vitro*; Bt, 244/55 CAZymes significantly expressed *in vivo/in vitro*) (Figures S2C and S2D). However, Bt H-2622 upregulates a higher proportion of GHs and PLs devoted to animal-derived carbohydrate utilization relative to Pc H-2477 (Figure 2E). Specifically, *in vivo* under high-MAC diet conditions, Bt H-2622 upregulates eight of 22 encoded mucus-targeted GHs (three out of 10 GH18; five out of 12 GH20), whereas Pc H-2477 encodes no GH18s and only one GH20, which is not upregulated in the high-MAC diet condition (Figures 2E, S2C, and S2D). In addition to targeting mucus carbohydrates, Bt H-2622 also

Table 1. Bacterial strains used

Strain name	Genus	Species	Strain	Origin specific	Origin	Genome size (bp)	Genome size (Mb)	Number of genes	Number of contigs
<i>Bt H-2209</i>	<i>Bacteroides</i>	<i>thetaitaomicron</i>	H-2209	Hadza	human feces	6,119,319	6.119319	4,722	1
<i>Bt H-2622</i>	<i>Bacteroides</i>	<i>thetaitaomicron</i>	H-2622	Hadza	human feces	6,037,034	6.037034	4,686	1
<i>Bc H-1617</i>	<i>Bacteroides</i>	<i>caccae</i>	H-1617	Hadza	human feces	5,112,756	5.112756	4,360	1
<i>Bf H-2631</i>	<i>Bacteroides</i>	<i>fragilis</i>	H-2631	Hadza	human feces	4,877,774	4.877774	4,622	3
<i>Bo H-1813</i>	<i>Bacteroides</i>	<i>ovatus</i>	H-1813	Hadza	human feces	6,715,646	6.715646	5,536	4
<i>Bo H-2495</i>	<i>Bacteroides</i>	<i>ovatus</i>	H-2495	Hadza	human feces	6,789,892	6.789892	5,086	1
<i>Bt VPI 5482</i>	<i>Bacteroides</i>	<i>thetaitaomicron</i>	VPI 5482	reference	human feces, unknown nationality	6,260,000	6.26	5,108	1
<i>Bc ATCC 43185</i>	<i>Bacteroides</i>	<i>caccae</i>	ATCC 43185	reference	human feces, Texas	5,280,000	5.28	4,399	1
<i>Bf NCTC 9343</i>	<i>Bacteroides</i>	<i>fragilis</i>	NCTC 9343	reference	human appendix abscess, UK	5,190,000	5.19	4,194	2
<i>Bo ATCC 8483</i>	<i>Bacteroides</i>	<i>ovatus</i>	ATCC 8483	reference	human feces, unknown nationality	6,470,000	6.47	4,896	1
<i>Pc DSM 18205</i>	<i>Prevotella</i>	<i>copri</i>	DSM 18205	reference	human feces, Japan	3,510,000	3.51	2,968	1
<i>Pc H-2379</i>	<i>Prevotella</i>	<i>copri</i>	H-2379	Hadza	human feces	4,350,632	4.350632	3,611	1
<i>Pc H-2383</i>	<i>Prevotella</i>	<i>copri</i>	H-2383	Hadza	human feces	4,506,031	4.506031	3,836	1
<i>Pc H-2446</i>	<i>Prevotella</i>	<i>copri</i>	H-2446	Hadza	human feces	4,057,255	4.057255	3,383	3
<i>Pc H-2477</i>	<i>Prevotella</i>	<i>copri</i>	H-2477	Hadza	human feces	4,111,062	4.111062	3,405	1
<i>Pc H-2489</i>	<i>Prevotella</i>	<i>copri</i>	H-2489	Hadza	human feces	4,115,122	4.115122	3,563	1
<i>Pc H-2497</i>	<i>Prevotella</i>	<i>copri</i>	H-2497	Hadza	human feces	4,081,238	4.081238	3,548	1
<i>Pc H-2632</i>	<i>Prevotella</i>	<i>copri</i>	H-2632	Hadza	human feces	3,850,424	3.850424	3,251	4
<i>Pc N-01</i>	<i>Prevotella</i>	<i>copri</i>	N-01	non-Hadza	human feces, USA	4,057,390	4.05739	3,880	1
<i>Pc YF2</i>	<i>Prevotella</i>	<i>copri</i>	YF2	reference	unknown	3,860,000	3.86	3,060	2

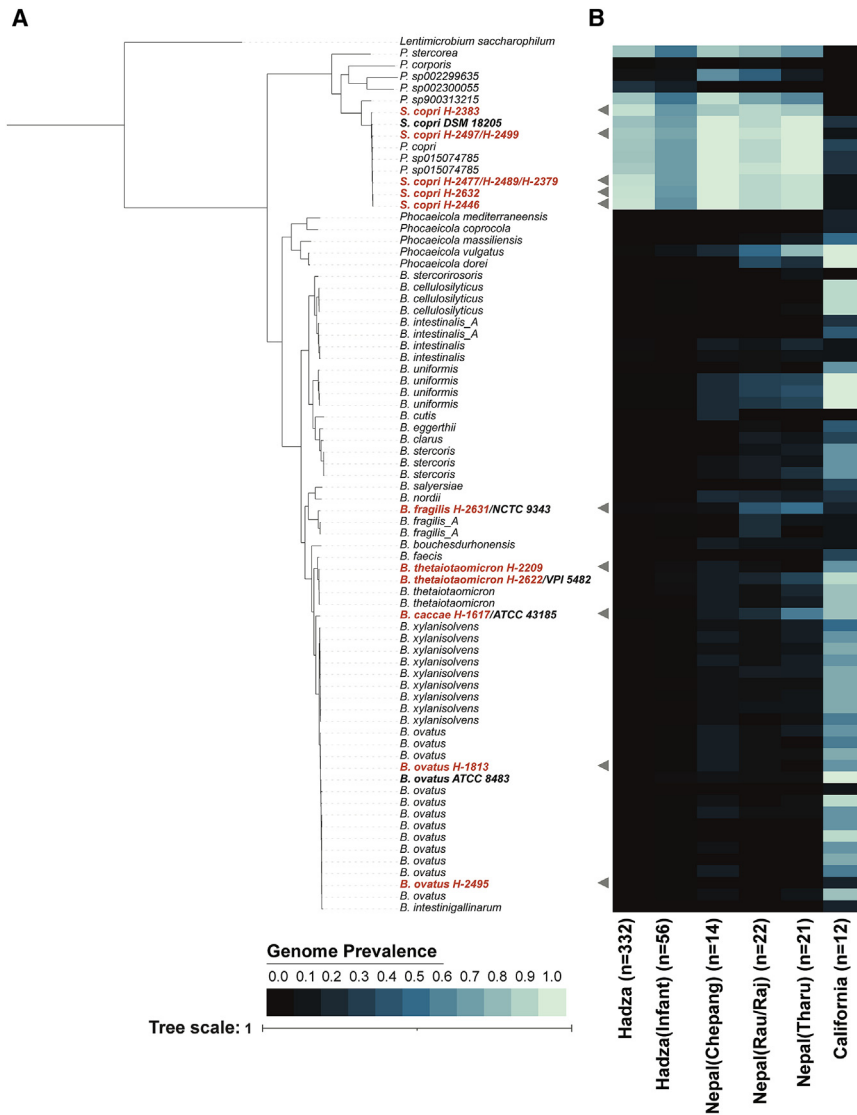


Figure 1. Hadza *Bacteroides* and *Prevotella* strains are related to previously sequenced isolates and vary in prevalence across populations

(A) Phylogenetic tree of *Prevotella* and *Bacteroides* genomes. Isolates from this study (red) and genomes from GenBank (black), strains used later in this study (bold).

(B) *Prevotella* and *Bacteroides* subspecies prevalence across foragers (Hadza and Chepang), agriculturalists (Rau, Raj, Tharu), or industrialized (California). Population size for each group is denoted as n. Prevalence defined as percentage of gut metagenomes from a population (column) in which a particular strain (row) is detected. Gray triangles indicate whole genomes isolated in this study, aligned with (A).

Taken together, these data indicate that, *in vivo*, *Bt H-2622* relies on both mucus and plant-derived carbohydrates. When plant carbohydrates are eliminated from the diet, *Bt H-2622* further upregulates mucus-degrading machinery, whereas *Pc H-2477*'s minimal mucus-degrading capacity renders it incapable of sustaining colonization in the absence of MACs.

Carbohydrate degradation capacity differences between Hadza *Bacteroides* and *Prevotella* mirrors industrialized strains

Hadza *Pc* and *Bacteroides* isolates have a similar number and predicted function of GHs and PLs to reference strains of the corresponding species (Table S3; Figure 3). Unsupervised clustering of GHs and PLs reveals that the Hadza strains cluster with their type strain counterparts, in keeping with the genetic similarity between genomes of the same species; the

upregulates 40 of its 97 plant-targeting GHs and PLs, whereas *Pc H-2477* upregulates all 38 of its plant-targeting GHs and PLs in the high-MAC diet (Figure 2E).

On the no-MAC diet relative to the *in vitro* condition, *Bt H-2622* upregulates two additional GH20s (along with the eight other mucin CAZymes upregulated on the high-MAC diet) as well as 27 plant-targeting GHs and PLs (Figures 2E and S2E). In other words, under high-MAC diet conditions, *Bt* upregulates CAZymes associated with plant, animal, and other carbohydrates equivalently (48 animal, 45 other, 46 plant), whereas, under the no-MAC diet condition, *Bt* upregulates a larger number and proportion of CAZymes associated with the utilization of animal associated carbohydrates (51 animal, 27 other, 35 plant). Since *Pc H-2477* does not colonize mice fed the no-MAC diet, this condition was not profiled. When comparing the no-MAC diet to the high-MAC diet, *Bt H-2622* differentially upregulates only three GHs, two of which degrade mucin (GH18) (Figure S2F).¹⁷

sets of CAZymes in each genome are most similar to the CAZymes found in genomes assigned to the same species (Figure 3A). When comparing the total number of GHs and PLs encoded within the Hadza strains to non-Hadza strains, we found similar total numbers of these genes and distribution of substrate specificity between strains of the same species (Figures 3B and 3C). Comparisons of Hadza *Pc* CAZymes are limited by the limited number of annotated *Pc* genomes in the CAZy database (only two exist at the time of this publication). We have now added seven more *Pc* genomes, and, as more *Pc* genomes are published, more variation in CAZyme repertoire may be uncovered.

While Hadza *Bacteroides* and *Prevotella* strains mirror the carbohydrate-degrading capacity of their non-Hadza counterparts, large differences exist between the *Bacteroides* and *Prevotella* strains. The *Bacteroides* encode more GHs and PLs than *Prevotella* strains even when corrected for genome size (251/21 average GH/PL in *Bacteroides*; 101/5 in *Prevotella*; Welch

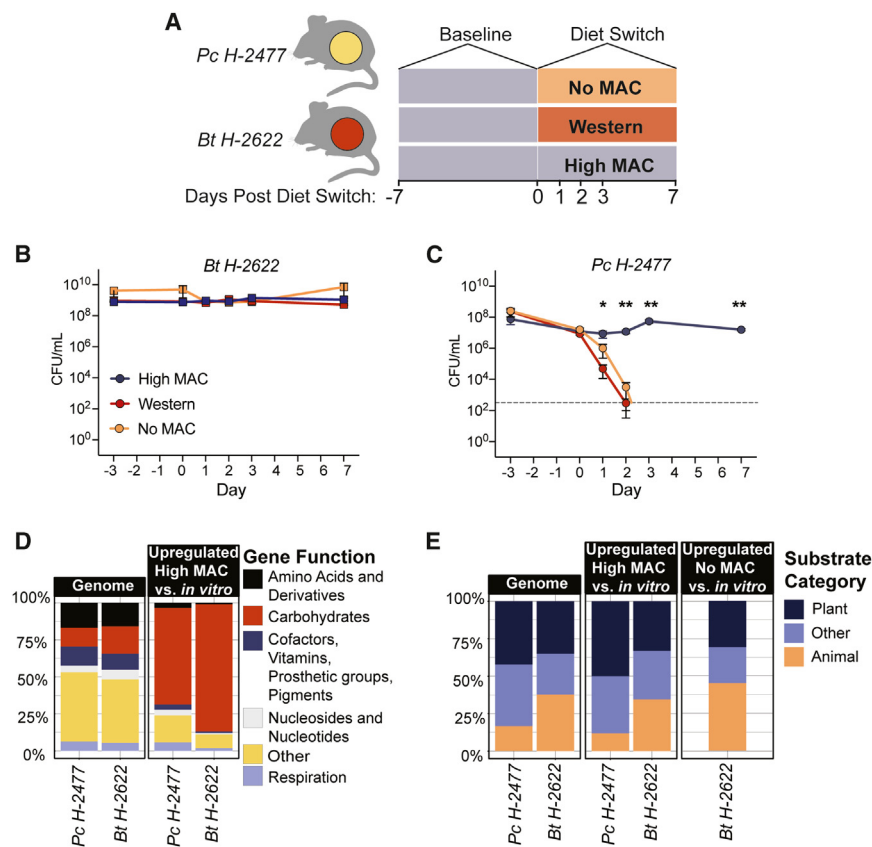


Figure 2. *Bt* and *Pc* colonization differ in diet-dependent manner

(A) Schematic of gnotobiotic experiments. (B and C) Fecal density of *Bt* H-2622 (B), and *Pc* H-2477 (C) in monocolonized mice ($n = 4/\text{group}$ for high-MAC and Western diets, $n = 5/\text{group}$ for no-MAC diet) fed different diets (multiple Mann-Whitney tests, $*p < 0.05$, $**p < 0.01$). Representative experiments, repeated twice. Dashed line denotes limit of detection (LOD = 500 CFU/mL). Error bars indicate standard error of the mean (SEM). (D) Proportion of genes upregulated *in vivo* in cecal contents of monocolonized mice on high-MAC diet on day seven of the experiment shown in (A), compared to gene expression in culture with PYG in late exponential phase. Genes organized by functional categories (Rapid Annotation using Subsystem Technology, RAST). (E) Proportion of predicted substrate categories of upregulated CAZymes in cecal contents of monocolonized under high-MAC conditions on day seven of the experiment shown in (A).

two-sample t test, $p = 0.0056$) (Table S3; Figure 3B). The proportion of *Bacteroides* GHs and PLs that are predicted to target plant carbohydrates or animal carbohydrates are equivalent (average 34% and 37%, respectively), whereas the *Prevotella*-encoded carbohydrate degradation is biased toward plant over animal carbohydrates (average 44% and 19%, respectively) (Figure 3C). The *Bacteroides* also encode a greater breadth of GH and PL families (averaging 68 CAZyme families per genome), while *Pc* isolates average 40 CAZy families per genome (Figure 3A), consistent with previously reported distributions for industrial-lifestyle-derived *Bacteroides* and *Prevotella* strains.³¹ The two genera also differ in their predicted mucin-degradation capacity (Figure S3; Wilcoxon test, $p = 3e-4$). CAZyme families GH18 and GH20 target carbohydrates found within the intestinal mucus lining.³⁷ All Hadza *Bacteroides* isolates harbor 11–14 GH20 and 1–13 GH18 CAZymes; however, the Hadza *Prevotella* isolates contain only one or two GH20s and only one isolate, *Pc* H-2497, contains a single GH18 (Figure 3D; Wilcoxon test, $p = 4e-4$).

The CAZyme contents of Hadza *Bacteroides* and *Prevotella* isolates are similar to their non-Hadza counterparts. Hadza *Bacteroides* isolates contain both more GHs and PLs overall as well as broader substrate-degrading capabilities that include both plant- and animal-derived carbohydrates relative to the Hadza *Prevotella* isolates. This difference between the Hadza *Bacteroides* and *Prevotella* strains is similar to that seen in non-Hadza strains and industrial lifestyle microbiotas,

suggesting that the *Prevotella* niche is more reliant upon plant carbohydrates compared to *Bacteroides*.^{38,39}

Dietary MACs are sufficient to maintain *Pc* colonization in the presence of *Bt*

To test whether Hadza *Bacteroides* and *Prevotella* isolates differ in their ability to use plant- and mucus-derived carbohydrates, we cultured Hadza and type strain *Bacteroides* and *Pc* isolates in media containing the plant carbohydrate inulin, porcine gastric mucin glycans, porcine intestinal heparin, or fructose as the sole carbon source. There is a range of ability to utilize inulin across the strains, consistent with previous work (Figure 4A).⁴⁰ Growth in the presence of mucin, however, is divided by genera; most *Bacteroides* isolates grow well on mucin, but the *P. copri* isolates do not (Figure 4A). These data are consistent with the lack of mucin-degrading capacity within the *Pc* genomes and the loss of *Pc* colonization *in vivo* when the host is the major carbohydrate source.

To determine whether the lack of diet-derived MACs is responsible for the loss of *Pc* H-2477 colonization we observed in the high-fat/low-MAC Western diet and no MAC diet (Figure 2C), we fed mice monocolonized with *Pc* H-2477 a high-MAC diet and then switched to either a custom diet containing 34% inulin by weight as the sole fermentable carbohydrate to match MAC content of the high-MAC diet (custom diets use gelatin as a binding agent and are noted by a “-g”; inulin-g) or a no-MAC diet (no-MAC-g).⁴¹ The no-MAC-g diet did not sustain *Pc* H-2477 colonization, with the strain becoming undetectable within 3 days (Figure 4B). However, *Pc* H-2477 maintained colonization in the presence of the inulin-g diet to levels similar to those observed in the high-MAC diet (Figures 2C and 4B), consistent with the requirement of MACs for *Pc* H-2477 colonization *in vivo*.

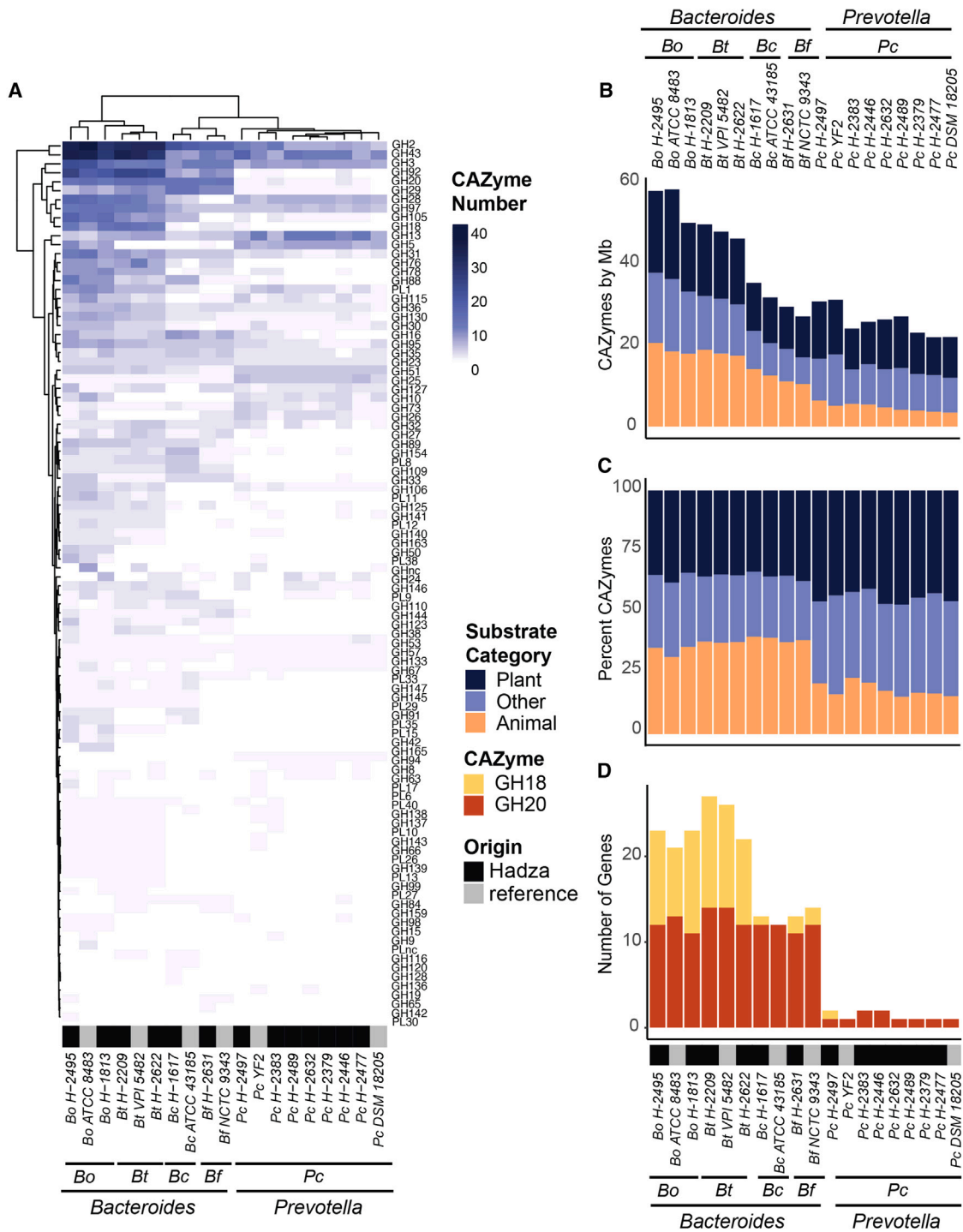


Figure 3. Hadza *Bacteroides* and *Prevotella* differ in distribution of GHs and PLs

(A) Number of GHs and PLs per genome indicated by CAZy family (rows). CAZymes shown appear at least once in any of the genomes analyzed. Hierarchical clustering via complete-linkage clustering method.

(B) Number of GHs and PLs normalized to genome size (Mb), colored by predicted substrate.

(C) Proportion of GHs and PLs in each genome colored by predicted substrate.

(D) Number of mucin-degrading GH18 and GH20 genes per genome.

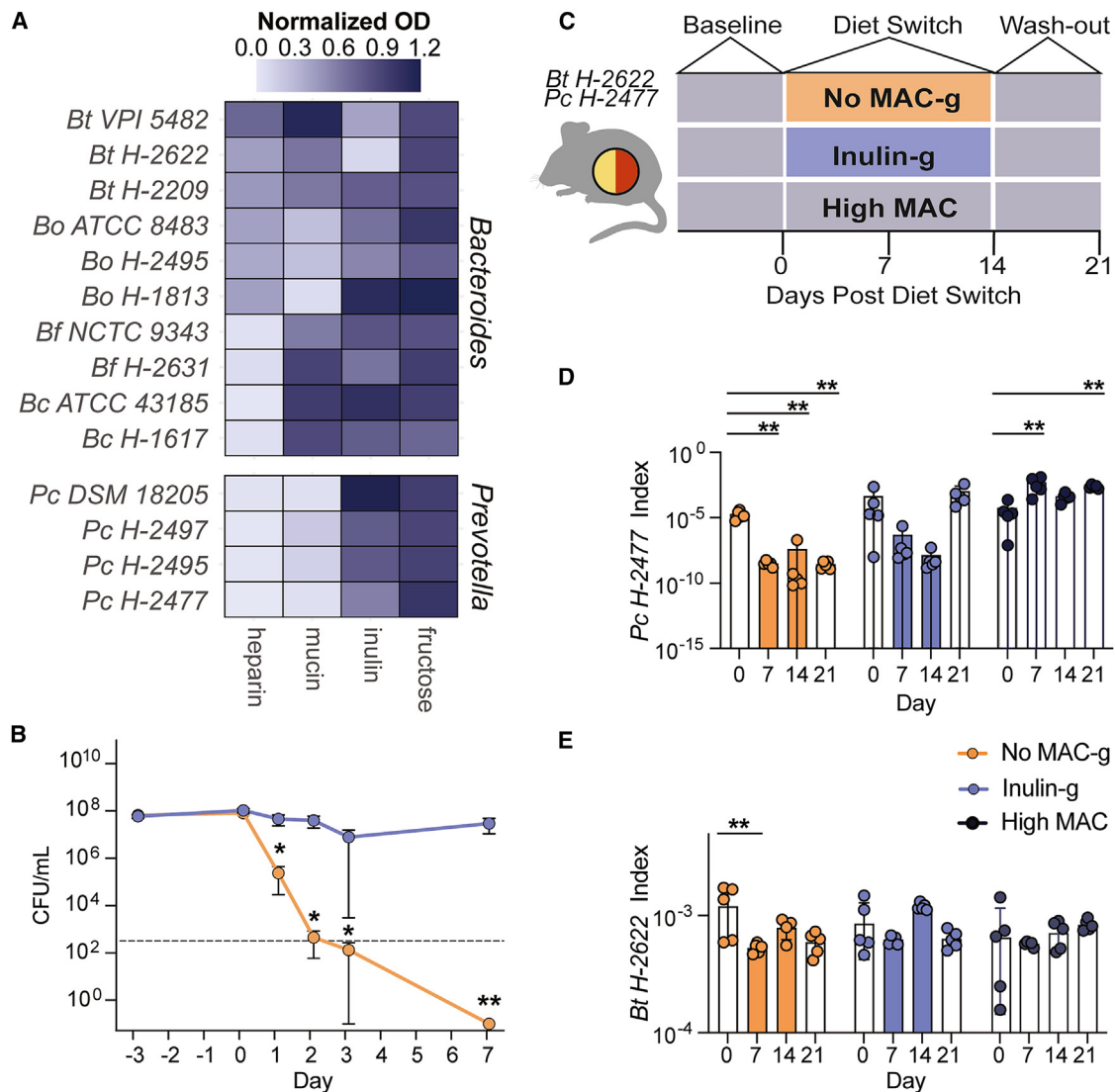


Figure 4. Reintroduction of MACs is sufficient to maintain *P. copri* colonization

(A) Normalized maximum optical density 600 (OD₆₀₀) of *Bacteroides* and *Prevotella* isolates grown in Yeast Casitone Fatty Acids broth (YCFA) with a single added carbohydrate for 24 h.

(B) Fecal CFUs of *Pc* H-2477 in monocolonized mice fed a no-MAC-g or inulin-g diet (mean + SEM, n = 5 mice per group, multiple Mann-Whitney tests, *p ≤ 0.05, **p < 0.01). Dashed line denotes LOD = 500 CFU/mL. Error bars indicate SEM. Representative experiment, repeated three times.

(C) Schematic of bicolonization with *Pc* H-2477 and *Bt* H-2622.

(D and E) qPCR index of DNA of *Pc* H-2477 (D) and *Bt* H-2622 (E) quantified from fecal samples from bicolonized mice (mean + SEM, n = 5 mice per group; multiple t tests, *p ≤ 0.05, **p < 0.01). Dashed line denotes LOD = 10⁻⁷ index. Error bars indicate SEM. Shaded bars indicate administration of high-MAC diet. Representative experiment, repeated twice.

We were curious how dietary MACs affect the relative abundance of *Pc* and *Bt* in mice when colonized together. GF mice were co-colonized with *Pc* H-2477 and *Bt* H-2622 and fed a high-MAC diet for 7 days and then either maintained on the high-MAC diet, switched to the no-MAC-g diet, or switched to the inulin-g diet for 2 weeks, followed by a 1-week period in which all mice consumed the high-MAC diet (Figure 4C). Prior to the diet switch (day 0), mice harbored both *Pc* H-2477 and *Bt* H-2622 with *Pc* abundance significantly lower than *Bt* (*Pc* index = 1.9e-4; *Bt* index = 9e-4; unpaired t test, p = 0.002; n = 15)

(Figures 4D and 4E). However, 7 days after the switch to the no-MAC-g diet, *Pc* H-2477 was no longer detectable, whereas *Bt* H-2622 colonization remained the same. The switch to the inulin-g diet resulted in a less dramatic decrease, with *Pc* still detectable after 7 days but not after 14 days, indicating that inulin provided support to *Pc* beyond the no-MAC diet (Figure 4D). *Bt* colonization remained stable on the inulin-g diet, with a small but significant increase in abundance on day 14 relative to the high-MAC condition (Figure 4E). When mice were returned to the high-MAC diet on day 14, those fed the inulin-g diet regained

relative abundance of *Pc H-2477* equivalent to that of baseline and to mice fed the high-MAC diet throughout the experiment. However, in mice switched to the high-MAC diet from the no-MAC-g diet, *Pc H-2477* DNA remained undetectable. *Bt* levels stayed constant in the no-MAC diet condition, with a small decrease on day 21 relative to high-MAC-fed mice (Figure 4E). These data are consistent with the requirement of dietary MACs for *Pc* colonization in the presence of *Bt* and indicate that the variety of carbohydrates in the high-MAC diet (derived from wheat, corn, oats, and alfalfa) better supports *Pc* colonization relative to a single-MAC diet (inulin). Given that *Bt*'s consumption of inulin is minimal *in vivo* (Figure 4A), it was unexpected that *Pc* colonization decreased in the inulin-g diet condition (Figure 4D). It is possible that *Pc* may face competition from *Bt* for the host-sourced carbohydrates it can access (Figure 2E) or that *Bt* colonization may alter the environment such that *Pc* abundance is affected in the inulin-g diet but not the high-MAC diet (Figure 2E). Our data further indicate that prolonged absence of MACs restricts the ability of *Pc* to regain abundance when dietary MACs are reintroduced.

DISCUSSION

The tradeoff between a microbiota dominated by *Bacteroides* or *Prevotella* based on host lifestyle has been well described, but its basis is not well understood.^{8,42}

Here, we demonstrate that Hadza isolates of *Bacteroides* and *Prevotella* do not differ dramatically from their non-Hadza counterparts in terms of genome-wide average nucleotide identity and carbohydrate utilization, suggesting that differences in their relative abundance and prevalence across lifestyle is not due to an inherent property of the population-specific strains themselves but to differences in their environments. Furthermore, we demonstrate that MACs are crucial for *Prevotella* to maintain colonization: even as the sole microbe, *Prevotella* is eradicated when dietary MACs are removed. *Bacteroides* species, however, can maintain colonization in the absence of dietary MACs due to their ability to use both plant- and host-derived carbohydrates, enabling continued colonization in low-MAC industrialized diets. Our data demonstrate that, in the presence of dietary MACs in gnotobiotic models, Hadza *Bacteroides* and *Prevotella* can co-exist, as is seen in the Hadza microbiota. However, removal of dietary MACs results in a precipitous decline in *Prevotella*, which does not recover when MACs are reintroduced. The presence of a single MAC, inulin, in the diet was sufficient to maintain an intermediate level of colonization that then rebounded when a more complete palate of MACs was available. These data are reminiscent of the seasonal pattern of *Prevotella* abundance in the Hadza, which cycles in abundance with the seasonality of their diet.

All together, these data are consistent with the model that, prior to industrialization, human microbiotas harbored both *Bacteroides* and *Prevotella* species. As diets shifted from high-MAC foraged foods to low-MAC industrially produced foods, abundance and prevalence of *Prevotella* diminished to the point of extinction in some individuals.⁴ *Prevotella* has been associated with beneficial health states, including improved glucose metabolism and increased resistance to malnutrition.^{43,44} However, *Prevotella* species have also been linked to negative outcomes,

including rheumatoid arthritis and insulin sensitivity.^{45,46} Strain-level variation, differences in other members of the microbiota (i.e., context-specific effects), and differences in host immune status could account for these contradictions. While a high-MAC diet is broadly beneficial to overall health, whether the presence of *Prevotella* affects these benefits is unknown.^{47–50} Additionally, how the loss of *Prevotella* and increased abundance of *Bacteroides* within the industrialized microbiota affects human physiology remains an important question.

Limitations of the study

Prior to this study, very few human-derived isolates of *Pc* were available for study, making comparisons of the strains isolated here to existing strains limited. As more isolates of *Pc* become available, it will be important to update the comparisons performed in this study. Additionally, while we demonstrate that *Pc* colonization decreases in response to a no-MAC diet, this colonization decline occurs so rapidly that we were not able to capture transcriptional data under this diet condition. This dataset would have been a useful comparison to *Pc* colonized mice on the high-MAC diet and *Bt*-colonized mice on the no-MAC diet. Furthermore, our study measured the effect of diet on *Pc* colonization, which was profound, but it is possible that there are other factors outside of diet that are important regulators of *Pc* colonization *in vivo*.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
 - Bacterial culture
 - Mouse husbandry
 - Statement on work with indigenous communities
- METHOD DETAILS
 - Strain isolation from fecal samples
 - Whole genome sequencing
 - Clustering genomes into subspecies
 - Evaluating subspecies prevalence and phylogenetic analysis
 - CAZyme annotation
 - *In vitro* polysaccharide growth assays
 - Colonization and enumeration of gnotobiotic mice
 - *In vivo* competition assays
 - Mouse diets
 - RNAseq
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.celrep.2023.113233>.

ACKNOWLEDGMENTS

We are indebted to the participants that provided samples used in this study. We acknowledge the numerous people and organizations who provided logistical support and conducted sample collection in the USA, Tanzania, and Nepal, including Dorobo Safaris, the Human Food Project, John Changalucha, Alphaxard Manjurano, Maria Gloria Domiguez-Bello, Allison Weakley, Samuel Smits, Gabriela Fragiadakis, Hannah Wastyk, Yoshina Gautam, Dinesh Bhandari, Sarmila Tandukar, Katharine Ng, Guru Prasad Gautam, Jeevan B. Sherehand, and members of the Gardner lab at Stanford. We thank Bryan Merrill, Madeline Topf, and Michelle St. Onge for technical support; Gabriela Gall Rosa for help with analysis; Samuel Lancaster, Brittany Sison, and Audrey Zhang for assistance processing RNA sequencing data; Michael Fischbach and Niokhor Dione for providing *P. copri N-01*; Brian Yu and Rose Yan for sequencing support; David Schneider for advice and tissue culture hood use; and Bernard Henrissat for advice on CAZyme annotation.

R.H.G. was supported by an NIH/NHGRI T32 training grant and the Blavatnik Family Foundation. This work was supported by grants from the NIH to J.L.S. (R01-DK085025, DP1-AT009892). J.L.S. is a Chan Zuckerberg Biohub investigator.

AUTHOR CONTRIBUTIONS

R.H.G., M.R.O., N.T., F.E., and S.K.H. performed the experiments and designed and performed data analyses. R.H.G., J.L.S., and E.D.S. conceived of the study and wrote the manuscript. J.L.S. and E.D.S. contributed to experimental design and data analysis and interpretation.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: March 28, 2023

Revised: August 16, 2023

Accepted: September 22, 2023

Published: October 28, 2023

REFERENCES

- Sonnenburg, J.L., and Sonnenburg, E.D. (2019). Vulnerability of the industrialized microbiota. *Science* 366, eaaw9255. <https://doi.org/10.1126/science.aaw9255>.
- De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J.B., Massart, S., Collini, S., Pieraccini, G., and Lionetti, P. (2010). 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. <https://doi.org/10.1073/pnas.1005963107>.
- Jha, A.R., Davenport, E.R., Gautam, Y., Bhandari, D., Tandukar, S., Ng, K.M., Fragiadakis, G.K., Holmes, S., Gautam, G.P., Leach, J., et al. (2018). Gut microbiome transition across a lifestyle gradient in Himalaya. *PLoS Biol.* 16, e2005396. <https://doi.org/10.1371/journal.pbio.2005396>.
- Carter, M.M., Olm, M.R., Merrill, B.D., Dahan, D., Tripathi, S., Spencer, S.P., Yu, F.B., Jain, S., Neff, N., Jha, A.R., et al. (2023). Ultra-deep sequencing of Hadza hunter-gatherers recovers vanishing gut microbes. *Cell* 186, 3111–3124.e13. <https://doi.org/10.1016/j.cell.2023.05.046>.
- Olm, M.R., Dahan, D., Carter, M.M., Merrill, B.D., Yu, F.B., Jain, S., Meng, X., Tripathi, S., Wastyk, H., Neff, N., et al. (2022). Robust variation in infant gut microbiome assembly across a spectrum of lifestyles. *Science* 376, 1220–1223. <https://doi.org/10.1126/science.abj2972>.
- Smits, S.A., Leach, J., Sonnenburg, E.D., Gonzalez, C.G., Lichtman, J.S., Reid, G., Knight, R., Manjurano, A., Changalucha, J., Elias, J.E., et al. (2017). Seasonal Cycling in the Gut Microbiome of the Hadza Hunter-Gatherers of Tanzania Authors. *Science* 357, 802–806. <https://doi.org/10.1126/science.aan4834>.
- Vangay, P., Johnson, A.J., Ward, T.L., Kashyap, P.C., Culhane-Pera, K.A., and Knights Correspondence, D. (2018). US Immigration Westernizes the Human Gut Microbiome. *Cell* 175, 962–972. <https://doi.org/10.1016/j.cell.2018.10.029>.
- Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P., et al. (2012). Human gut microbiome viewed across age and geography. *Nature* 486, 222–227. <https://doi.org/10.1038/nature11053>.
- Wibowo, M.C., Yang, Z., Borry, M., Hübner, A., Huang, K.D., Tierney, B.T., Zimmerman, S., Barajas-Olmos, F., Contreras-Cubas, C., García-Ortiz, H., et al. (2021). Reconstruction of ancient microbial genomes from the human gut. *Nature* 594, 234–239. <https://doi.org/10.1038/s41586-021-03532-0>.
- Sonnenburg, E.D., and Sonnenburg, J.L. (2014). Starving our microbial self: The deleterious consequences of a diet deficient in microbiota-accessible carbohydrates. *Cell Metab.* 20, 779–786. <https://doi.org/10.1016/j.cmet.2014.07.003>.
- Cordain, L., Eaton, S.B., Sebastian, A., Mann, N., Lindeberg, S., Watkins, B.A., O'Keefe, J.H., and Brand-Miller, J. (2005). Origins and evolution of the Western diet: health implications for the 21st century. *Am. J. Clin. Nutr.* 81, 341–354. <https://doi.org/10.1093/ajcn.81.2.341>.
- Flint, H.J., Scott, K.P., Duncan, S.H., Louis, P., and Forano, E. (2012). Microbial degradation of complex carbohydrates in the gut. *Gut Microb.* 3, 289–306. <https://doi.org/10.4161/GMIC.19897>.
- Bell, A., and Juge, N. (2021). Mucosal glycan degradation of the host by the gut microbiota. *Glycobiology* 31, 691–696. <https://doi.org/10.1093/GLYCOB/CWAA097>.
- Desai, M.S., Seekatz, A.M., Koropatkin, N.M., Kamada, N., Hickey, C.A., Wolter, M., Pudlo, N.A., Kitamoto, S., Terrapon, N., Muller, A., et al. (2016). A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. *Cell* 167, 1339–1353.e21. <https://doi.org/10.1016/J.CELL.2016.10.043>.
- Pudlo, N.A., Urs, K., Crawford, R., Pirani, A., Atherly, T., Jimenez, R., Terrapon, N., Henrissat, B., Peterson, D., Ziemer, C., et al. (2022). Phenotypic and Genomic Diversification in Complex Carbohydrate-Degrading Human Gut Bacteria. *mSystems* 7, e0094721. <https://doi.org/10.1128/MSYS-TEMS.00947-21>.
- Salyers, A.A., West, S.E., Vercellotti, J.R., and Wilkins, T.D. (1977). Fermentation of mucins and plant polysaccharides by anaerobic bacteria from the human colon. *Appl. Environ. Microbiol.* 34, 529–533. <https://doi.org/10.1128/aem.34.5.529-533.1977>.
- Sonnenburg, J.L., Xu, J., Leip, D.D., Chen, C.H., Westover, B.P., Weatherford, J., Buhler, J.D., and Gordon, J.I. (2005). Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science* 307, 1955–1959. https://doi.org/10.1126/SCIENCE.1109051/SUPPL_FILE/SONNENBURG.SOM.PDF.
- Earle, K.A., Billings, G., Sigal, M., Lichtman, J.S., Hansson, G.C., Elias, J.E., Amieva, M.R., Huang, K.C., and Sonnenburg, J.L. (2015). Quantitative Imaging of Gut Microbiota Spatial Organization. *Cell Host Microbe* 18, 478–488. <https://doi.org/10.1016/j.chom.2015.09.002>.
- Martens, E.C., Neumann, M., and Desai, M.S. (2018). Interactions of commensal and pathogenic microorganisms with the intestinal mucosal barrier. *Nat. Rev. Microbiol.* 16, 457–470. <https://doi.org/10.1038/s41579-018-0036-x>.
- Sonnenburg, E.D., Smits, S.A., Tikhonov, M., Higginbottom, S.K., Wingreen, N.S., and Sonnenburg, J.L. (2016). Diet-induced extinctions in the gut microbiota compound over generations. *Nature* 529, 212–215. <https://doi.org/10.1038/nature16504>.
- Kaplan, R.C., Wang, Z., Usyk, M., Sotres-Alvarez, D., Daviglius, M.L., Schneiderman, N., Talavera, G.A., Gellman, M.D., Thyagarajan, B., Moon, J.-Y., et al. (2019). Gut microbiome composition in the Hispanic Community Health Study/Study of Latinos is shaped by geographic

- relocation, environmental factors, and obesity. *Genome Biol.* 20, 219. <https://doi.org/10.1186/s13059-019-1831-z>.
22. Zafar, H., and Saier, M.H. (2021). Gut *Bacteroides* species in health and disease. *Gut Microb.* 13, 1–20. <https://doi.org/10.1080/19490976.2020.1848158>.
 23. Tett, A., Pasolli, E., Masetti, G., Ercolini, D., and Segata, N. (2021). Prevotella diversity, niches and interactions with the human host. *Nat. Rev. Microbiol.* 19, 585–599. <https://doi.org/10.1038/s41579-021-00559-y>.
 24. Li, H., Meier-Kolthoff, J.P., Hu, C., Wang, Z., Zhu, J., Zheng, W., Tian, Y., and Guo, F. (2022). Panoramic Insights into Microevolution and Macroevolution of A Prevotella copri-containing Lineage in Primate Guts. *Dev. Reprod. Biol.* 20, 334–349. <https://doi.org/10.1016/j.gpb.2021.10.006>.
 25. Abdill, R.J., Adamowicz, E.M., and Blekhman, R. (2022). Public human microbiome data are dominated by highly developed countries. *PLoS Biol.* 20, e3001536. <https://doi.org/10.1371/journal.pbio.3001536>.
 26. Accetto, T., and Avguštin, G. (2015). Polysaccharide utilization locus and CAZYme genome repertoires reveal diverse ecological adaptation of Prevotella species. *Syst. Appl. Microbiol.* 38, 453–461. <https://doi.org/10.1016/J.SYAPM.2015.07.007>.
 27. Li, J., Gálvez, E.J.C., Amend, L., Almási, É., Iljazovic, A., Lesker, T.R., Bielecka, A.A., Schorr, E.-M., and Strowig, T. (2021). A versatile genetic toolbox for Prevotella copri enables studying polysaccharide utilization systems. *EMBO J.* 40, e108287. <https://doi.org/10.15252/embj.2021108287>.
 28. Xu, J., Bjursell, M.K., Himrod, J., Deng, S., Carmichael, L.K., Chiang, H.C., Hooper, L.V., and Gordon, J.I. (2003). A Genomic View of the Human-Bacteroides thetaiotaomicron Symbiosis. *Science* 299, 2074–2076. <https://doi.org/10.1126/science.1080029>.
 29. Bjursell, M.K., Martens, E.C., and Gordon, J.I. (2006). Functional Genomic and Metabolic Studies of the Adaptations of a Prominent Adult Human Gut Symbiont, Bacteroides thetaiotaomicron, to the Suckling Period * □ S Downloaded from. *J. Biol. Chem.* 281, 36269–36279. <https://doi.org/10.1074/jbc.M606509200>.
 30. Dodd, D., Moon, Y.-H., Swaminathan, K., Mackie, R.I., and Cann, I.K.O. (2010). Transcriptomic analyses of xylan degradation by Prevotella bryantii and insights into energy acquisition by xylanolytic bacteroidetes. *J. Biol. Chem.* 285, 30261–30273. <https://doi.org/10.1074/jbc.M110.141788>.
 31. Fehlner-Peach, H., Magnabosco, C., Raghavan, V., Scher, J.U., Tett, A., Cox, L.M., Gottsegen, C., Watters, A., Wiltshire-Gordon, J.D., Segata, N., et al. (2019). Distinct Polysaccharide Utilization Profiles of Human Intestinal Prevotella copri Isolates. *Cell Host Microbe* 26, 680–690.e5. <https://doi.org/10.1016/j.chom.2019.10.013>.
 32. Sprockett, D.D., Martin, M., Costello, E.K., Burns, A.R., Holmes, S.P., Gurven, M.D., and Relman, D.A. (2020). Microbiota assembly, structure, and dynamics among Tsimane horticulturalists of the Bolivian Amazon. *Nat. Commun.* 11, 3772. <https://doi.org/10.1038/s41467-020-17541-6>.
 33. De Filippis, F., Pasolli, E., Tett, A., Tarallo, S., Naccarati, A., De Angelis, M., Neviani, E., Cocolin, L., Gobbetti, M., Segata, N., and Ercolini, D. (2019). Distinct Genetic and Functional Traits of Human Intestinal Prevotella copri Strains Are Associated with Different Habitual Diets. *Cell Host Microbe* 25, 444–453.e3. <https://doi.org/10.1016/j.chom.2019.01.004>.
 34. Tett, A., Huang, K.D., Asnicar, F., Fehlner-Peach, H., Pasolli, E., Karcher, N., Armanini, F., Manghi, P., Bonham, K., Zolfo, M., et al. (2019). The Prevotella copri Complex Comprises Four Distinct Clades Underrepresented in Westernized Populations. *Cell Host Microbe* 26, 666–679.e7. <https://doi.org/10.1016/j.chom.2019.08.018>.
 35. Marlowe, F.W., and Berbesque, J.C. (2009). Tubers as fallback foods and their impact on Hadza hunter-gatherers. *Am. J. Phys. Anthropol.* 140, 751–758. <https://doi.org/10.1002/ajpa.21040>.
 36. Monteiro, C.A., Moubarac, J.-C., Cannon, G., Ng, S.W., and Popkin, B. (2013). Ultra-processed products are becoming dominant in the global food system. *Obes. Rev.* 14, 21–28. <https://doi.org/10.1111/obr.12107>.
 37. Luis, A.S., Jin, C., Pereira, G.V., Glowacki, R.W.P., Gugel, S.R., Singh, S., Byrne, D.P., Pudlo, N.A., London, J.A., Baslé, A., et al. (2021). A single sulfatase is required to access colonic mucin by a gut bacterium. *Nat* 598, 332–337. <https://doi.org/10.1038/s41586-021-03967-5>.
 38. Gálvez, E.J.C., Iljazovic, A., Amend, L., Lesker, T.R., Renault, T., Thiemann, S., Hao, L., Roy, U., Gronow, A., Charpentier, E., et al. (2020). Distinct Polysaccharide Utilization Determines Interspecies Competition between Intestinal Prevotella spp. *Cell Host Microbe* 28, 838–852. <https://doi.org/10.1016/j.chom.2020.09.012>.
 39. Aakko, J., Pietilä, S., Toivonen, R., Rokka, A., Makkala, K., Laitinen, K., Elo, L., and Hänninen, A. (2020). A carbohydrate-active enzyme (CAZY) profile links successful metabolic specialization of Prevotella to its abundance in gut microbiota. *Sci. Rep.* 10, 12411. <https://doi.org/10.1038/s41598-020-69241-2>.
 40. Sonnenburg, E.D., Zheng, H., Joglekar, P., Higginbottom, S.K., Firkbank, S.J., Bolam, D.N., and Sonnenburg, J.L. (2010). Specificity of Polysaccharide Use in Intestinal Bacteroides Species Determines Diet-Induced Microbiota Alterations. *Cell* 141, 1241–1252. <https://doi.org/10.1016/j.cell.2010.05.005>.
 41. Dubos, R.J., and Pierce, C. (1948). The effect of diet on experimental tuberculosis of mice. *Am. Rev. Tuberc.* 57, 287–293.
 42. Gorvitovskaia, A., Holmes, S.P., and Huse, S.M. (2016). Interpreting Prevotella and Bacteroides as biomarkers of diet and lifestyle. *Microbiome* 4, 15. <https://doi.org/10.1186/s40168-016-0160-7>.
 43. Kovatcheva-Datchary, P., Nilsson, A., Akrami, R., Lee, Y.S., De Vadder, F., Arora, T., Hallen, A., Martens, E., Björck, I., and Bäckhed, F. (2015). Dietary Fiber-Induced Improvement in Glucose Metabolism Is Associated with Increased Abundance of Prevotella. *Cell Metab.* 22, 971–982. <https://doi.org/10.1016/J.CMET.2015.10.001>.
 44. Raman, A.S., Gehrig, J.L., Venkatesh, S., Chang, H.-W., Hibberd, M.C., Subramanian, S., Kang, G., Bessong, P.O., Lima, A.A.M., Kosek, M.N., et al. (2019). A sparse covarying unit that describes healthy and impaired human gut microbiota development. *Science* 365, eaau4735. <https://doi.org/10.1126/science.aau4735>.
 45. Scher, J.U., Sczesnak, A., Longman, R.S., Segata, N., Ubeda, C., Bielski, C., Rostron, T., Cerundolo, V., Pamer, E.G., Abramson, S.B., et al. (2013). Expansion of intestinal Prevotella copri correlates with enhanced susceptibility to arthritis. *Elife* 2, 1202. <https://doi.org/10.7554/eLife.01202.001>.
 46. Pedersen, H.K., Gudmundsdottir, V., Nielsen, H.B., Hyötyläinen, T., Nielsen, T., Jensen, B.A.H., Forsslund, K., Hildebrand, F., Prifti, E., Falony, G., et al. (2016). Human gut microbes impact host serum metabolome and insulin sensitivity. *Nature* 535, 376–381. <https://doi.org/10.1038/nature18646>.
 47. Lancaster, S.M., Lee-McMullen, B., Abbott, C.W., Quijada, J.V., Hornburg, D., Park, H., Perelman, D., Peterson, D.J., Tang, M., Robinson, A., et al. (2022). Global, distinctive, and personal changes in molecular and microbial profiles by specific fibers in humans. *Cell Host Microbe* 30, 848–862.e7. <https://doi.org/10.1016/j.chom.2022.03.036>.
 48. Ng, K.M., Aranda-Díaz, A., Tropini, C., Frankel, M.R., Van Treuren, W., O’Loughlin, C.T., Merrill, B.D., Yu, F.B., Pruss, K.M., Oliveira, R.A., et al. (2019). Recovery of the Gut Microbiota after Antibiotics Depends on Host Diet, Community Context, and Environmental Reservoirs. *Cell Host Microbe* 26, 650–665.e4. <https://doi.org/10.1016/j.chom.2019.10.011>.
 49. Yang, Y., Zhao, L.-G., Wu, Q.-J., Ma, X., and Xiang, Y.-B. (2015). Association between dietary fiber and lower risk of all-cause mortality: a meta-analysis of cohort studies. *Am. J. Epidemiol.* 181, 83–91. <https://doi.org/10.1093/aje/kwu257>.
 50. Kim, Y., and Je, Y. (2014). Dietary fiber intake and total mortality: a meta-analysis of prospective cohort studies. *Am. J. Epidemiol.* 180, 565–573. <https://doi.org/10.1093/aje/kwu174>.
 51. Parada, A.E., Needham, D.M., and Fuhrman, J.A. (2016). Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* 18, 1403–1414. <https://doi.org/10.1111/1462-2920.13023>.

52. Kolmogorov, M., Armstrong, J., Raney, B.J., Streeter, I., Dunn, M., Yang, F., Odom, D., Flicek, P., Keane, T.M., Thybert, D., et al. (2018). Chromosome assembly of large and complex genomes using multiple references. *Genome Res.* **28**, 1720–1732. <https://doi.org/10.1101/gr.236273.118>.
53. Brettin, T., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Olsen, G.J., Olson, R., Overbeek, R., Parrello, B., Pusch, G.D., et al. (2015). RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci. Rep.* **5**, 8365. <https://doi.org/10.1038/srep08365>.
54. Olm, M.R., Brown, C.T., Brooks, B., and Banfield, J.F. (2017). dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* **11**, 2864–2868. <https://doi.org/10.1038/ismej.2017.126>.
55. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359. <https://doi.org/10.1038/nmeth.1923>.
56. Olm, M.R., Crits-Christoph, A., Bouma-Gregson, K., Firek, B.A., Morowitz, M.J., and Banfield, J.F. (2021). inStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. *Nat. Biotechnol.* **39**, 727–736. <https://doi.org/10.1038/s41587-020-00797-0>.
57. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>.
58. Lee, M.D. (2019). GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics* **35**, 4162–4164. <https://doi.org/10.1093/bioinformatics/btz188>.
59. Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P., and Parks, D.H. (2019). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>.
60. Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296. <https://doi.org/10.1093/nar/gkab301>.
61. Drula, E., Garron, M.-L., Dogan, S., Lombard, V., Henrissat, B., and Terrapon, N. (2022). The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res.* **50**, D571–D577. <https://doi.org/10.1093/nar/gkab1045>.
62. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. *BMC Bioinf.* **10**, 421. <https://doi.org/10.1186/1471-2105-10-421>.
63. Mistry, J., Finn, R.D., Eddy, S.R., Bateman, A., and Punta, M. (2013). Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* **41**, e121. <https://doi.org/10.1093/nar/gkt263>.
64. Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048. <https://doi.org/10.1093/bioinformatics/btw354>.
65. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
66. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915. <https://doi.org/10.1038/s41587-019-0201-4>.
67. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008. <https://doi.org/10.1093/gigascience/giab008>.
68. Pertea, M., Kim, D., Pertea, G.M., Leek, J.T., and Salzberg, S.L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**, 1650–1667. <https://doi.org/10.1038/nprot.2016.095>.
69. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
70. Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., et al. (2019). Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686. <https://doi.org/10.21105/joss.01686>.
71. Wickham, H. (2022). *stringr: Simple, Consistent Wrappers for Common String Operations*.
72. Wickham, H., and Bryan, J. (2023). *readxl: Read Excel Files*.
73. Wickham, H. (2023). An SVG Graphics Device. <https://svglite-r-lib.org/>.
74. Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). circlize implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812. <https://doi.org/10.1093/bioinformatics/btu393>.
75. Gu, Z. (2022). Complex heatmap visualization. *iMeta* **1**, e43. <https://doi.org/10.1002/imt2.43>.
76. Sakai, R., Winand, R., Verbeiren, T., Moere, A.V., and Aerts, J. (2014). dendsort: modular leaf ordering methods for dendrogram representations in R. *F1000Research* **3**. <https://doi.org/10.12688/f1000research.4784.1>.
77. Wickham, H. (2016). Programming with ggplot2. In *ggplot2: Elegant Graphics for Data Analysis Use R!*, H. Wickham, ed. (Springer International Publishing), pp. 241–253. https://doi.org/10.1007/978-3-319-24277-4_12.
78. Hahsler, M., Hornik, K., and Buchta, C. (2008). Getting Things in Order: An Introduction to the R Package seriation. *J. Stat. Softw.* **25**, 1–34. <https://doi.org/10.18637/jss.v025.i03>.
79. Pruss, K.M., and Sonnenburg, J.L. (2021). *C. difficile* exploits a host metabolite produced during toxin-mediated disease. *Nat* **593**, 261–265. <https://doi.org/10.1038/s41586-021-03502-6>.
80. Green, E.D., Gunter, C., Biesecker, L.G., Di Francesco, V., Easter, C.L., Feingold, E.A., Felsenfeld, A.L., Kaufman, D.J., Ostrander, E.A., Pavan, W.J., et al. (2020). Strategic vision for improving human health at The Forefront of Genomics. *Nature* **586**, 683–692. <https://doi.org/10.1038/s41586-020-2817-4>.
81. Fragiadakis, G.K., Smits, S.A., Sonnenburg, E.D., Van Treuren, W., Reid, G., Knight, R., Manjuran, A., Chandalucha, J., Dominguez-Bello, M.G., Leach, J., and Sonnenburg, J.L. (2019). Links between environment, diet, and the hunter-gatherer microbiome. *Gut Microb.* **10**, 216–227. <https://doi.org/10.1080/19490976.2018.1494103>.
82. Johnson, M., Zaretskaya, I., Raytselis, Y., Merezuk, Y., McGinnis, S., and Madden, T.L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Res.* **36**, W5–W9. <https://doi.org/10.1093/nar/gkn201>.
83. Pribelski, A., Antipov, D., Meleshko, D., Lapidus, A., and Korobeynikov, A. (2020). Using SPAdes De Novo Assembler. *Curr. Protoc. Bioinforma.* **70**, e102. <https://doi.org/10.1002/cpbi.102>.
84. Mikheenko, A., Pribelski, A., Saveliev, V., Antipov, D., and Gurevich, A. (2018). Versatile genome assembly evaluation with QUASt-LG. *Bioinformatics* **34**, i142–i150. <https://doi.org/10.1093/bioinformatics/bty266>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Human fecal samples from Hadza people	Smits et al., 2017 ⁶	N/A
Critical commercial assays		
MasterPure Gram Positive DNA Purification Kit	LGC Biosearch Technologies	Cat#NC9197506
QIAquick PCR Purification Kit	Qiagen	Cat#28106
MinION Flowcell FLO-MIN106	Nanopore	Kit: SQK-LSK109 Barcode kit: EXP-NBD104
MiSeq	Illumina	https://www.illumina.com/systems/sequencing-platforms/miseq.html
Epoch2 Microplate Reader	BioTek	https://www.biotek.com/products/detection-microplate-readers/epoch-2-microplate-spectrophotometer/
DNeasy PowerLyzer PowerSoil	Qiagen	Cat#12855-50
Brilliant III, Ultra Fast SYBR Green QPCR Master Mix	Agilent	Cat#600883
CFX Connect Real-Time PCR Detection System	Bio-Rad	Cat#1855201
RNeasy PowerMicrobiome Kit	Qiagen	Cat#26000-50
RiboMinus™ Transcriptome Isolation Kit, bacteria	Invitrogen	Cat#K155004
TruSeq® Stranded Total RNA Library Prep Human/Mouse/Rat	Illumina	Cat#20020597
NovaSeq SP Flow Cell	Illumina	NovaSeq 6000 System
Deposited data		
Hadza 16S sequencing	Smits et al., 2017 ⁶	N/A
Metagenomic reads: Hadza, Nepal, and California populations	Carter et al., 2023 ⁴	N/A
Whole genome sequences: <i>Bacteroides</i> and <i>Prevotella</i> isolates	this study	NCBI BioProject PRJNA1015720: http://www.ncbi.nlm.nih.gov/bioproject/1015720
RNAseq data	this study	Zenodo: https://doi.org/10.5281/zenodo.7651179
<i>Bacteroides</i> and <i>Prevotella</i> reference genomes	GenBank	NCBI
Experimental models: Organisms/strains		
<i>P. copri</i> isolates	this study	N/A
<i>Bacteroides</i> sp. isolates	this study	N/A
<i>Prevotella copri</i> DSM 18205	DSMZ	Cat#DSM 18205
<i>Bacteroides ovatus</i> ATCC 8483	ATCC	Cat#8483
<i>Bacteroides thetaiotaomicron</i> VPI-5482	ATCC	Cat#29148
<i>Bacteroides fragilis</i> NCTC 9343	ATCC	Cat#25285
<i>Bacteroides caccae</i> ATCC 43185	ATCC	Cat#43185
Mouse: Swiss Webster, germ free	Taconic	Cat#SW-F GF
Oligonucleotides		
<i>P. copri</i> forward primer: hpc_gyrB_03_F1	this study	CACCCACACCATGTAAACCGCCAG
<i>P. copri</i> reverse primer: hpc_gyrB_03_R	this study	TGTACCGACATCGAAGTTACCATCAACGAAG
<i>B. thetaiotaomicron</i> forward primer: HBT05_03F	this study	GCAGGCACGGGCAGTATCAGTATCG
<i>B. thetaiotaomicron</i> reverse primer: HBT05_03R	this study	CGCCACGGATAGGCAGACATTTGTCA
Bacterial 16S forward primer: 16S rRNA 515F	Parada et al., 1998 ⁵¹	5'-GTGYCAGCMGCCGCGGTAA-3'
Bacterial 16S reverse primer: 16S rRNA 806R	Parada et al., 1998 ⁵¹	5'-GGACTACHVGGGTWTCTAAT-3'

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
FastQC	Babraham Bioinformatics	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
BBTools	Joint Genome Institute	https://jgi.doe.gov/data-and-tools/software-tools/bbtools/
SPAdes	Center for Algorithmic Biotechnology	https://cab.spbu.ru/software/spades/
RagOUT	Kolmogorov et al., 2018 ⁵²	https://github.com/fenderglass/Ragout
Quast	Center for Algorithmic Biotechnology	https://quast.sourceforge.net/index.html
RASTtk	Brettin et al., 2015 ⁵³	https://rast.nmpdr.org/
dRep(v3.2.1)	Olm et al., 2017 ⁵⁴	https://github.com/MrOlm/drep
Bowtie2	Langmeade and Salzberg, 2012 ⁵⁵	https://bowtie-bio.sourceforge.net/bowtie2/index.shtml
inStrain	Olm et al., 2021 ⁵⁶	https://github.com/MrOlm/inStrain
scipy.cluster.hierarchy	Virtanen et al., 2020 ⁵⁷	https://docs.scipy.org/doc/scipy/reference/cluster.hierarchy.html
GToTree (version 1.5.36)	Lee, MD, 2019 ⁵⁸	https://github.com/AstroBioMike/GToTree/tree/v1.5.36
GTDB	Chaumeil et al., 2020 ⁵⁹	https://gtdb.ecogenomic.org/
iTol	Letunic and Bork, 2021 ⁶⁰	https://itol.embl.de/
CAZy	Drula et al., 2022 ⁶¹	http://www.cazy.org/
BlastP (version 2.3.0+)	Camacho et al., 2009 ⁶²	https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE = BlastDocs&DOC_TYPE = Download
HMMER (version 3.1)	Mistry et al., 2013 ⁶³	http://hmmer.org/download.html
Multiqc (version 1.14)	Ewels et al., 2016 ⁶⁴	https://multiqc.info/
Trimmomatic (version 0.39)	Bolger et al., 2014 ⁶⁵	http://www.usadellab.org/cms/?page = trimmomatic
HiSAT2 (version 2.2.0)	Kim et al., 2019 ⁶⁶	http://daehwankimlab.github.io/hisat2/
SAMtools (version 1.16.1)	Danecek et al., 2021 ⁶⁷	http://www.htslib.org/
StringTie (version 2.1.3)	Shumate et al., 2022 ⁶⁸	http://ccb.jhu.edu/software/stringtie/index.shtml?t = manual
DESeq2 (version 1.38.3)	Love et al., 2014 ⁶⁹	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
R (version 4.2.2)	R Core Team	https://www.r-project.org/
tidyverse (version 1.3.2)	Wickham et al., 2019 ⁷⁰	https://www.tidyverse.org/
RStudio (version 1.4)	R Core Team	https://www.rstudio.com/
stringr (version 1.5.0)	Wickham, 2022 ⁷¹	https://stringr.tidyverse.org
MetBrewer (version 0.2.0)	Blake Mills	https://github.com/BlakeRMills/MetBrewer
RColorBrewer (version 1.1–3)	Erich Neuwirth	https://cran.r-project.org/web/packages/RColorBrewer/index.html
cowplot (version 1.1.1)	Claus Wilke	https://cran.r-project.org/web/packages/cowplot/vignettes/introduction.html
readxl (version 1.4.1)	Wickham and Bryan, 2023 ⁷²	https://readxl.tidyverse.org/
svglite (version 2.1.1)	Wickham, 2023 ⁷³	https://svglite.r-lib.org/
circize (version 0.4.15)	Gu et al., 2014 ⁷⁴	https://jokergoo.github.io/circize_book/book/
ComplexHeatmap (version 2.14.0)	Gu, 2022 ⁷⁵	https://jokergoo.github.io/ComplexHeatmap-reference/book/
dendsort version 0.3.4	Sakai et al., 2014 ⁷⁶	https://cran.rstudio.com/web/packages/dendsort/index.html
dendextend version 1.16.0	Tal Galili	https://www.rdocumentation.org/packages/dendextend/versions/1.16.0#how-to-cite-the-dendextend-package

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
ggplot2 version 3.4.0	Wickham, 2016 ⁷⁷	https://cran.r-project.org/web/packages/ggplot2/index.html
seriation version 1.4.1	Hahsler et al., 2008 ⁷⁸	https://www.rdocumentation.org/packages/seriation/versions/1.4.1
Prism 9 for macOS	GraphPad Software	https://www.graphpad.com/guides/prism/latest/user-guide/citing_graphpad_prism.htm

Other

Heparin sodium salt (from porcine intestinal mucosa)	Sigma	H3393-50KU
D-Glucose, Anhydrous	Alfa-aesar	aaa16828-0e
D-Fructose 99%	Alfa-aesar	AAA17718-30
Inulin (chicory)	Beneo	Orafti®HP
KAIBAE Premium Baobab Fruit Powder	Amazon	N/A
Mucin from porcine stomach, Type III, bound sialic acid 0.5–1.5%, partially purified powder	Sigma	M1778-100G
<i>Bacteroides</i> Bile Esculin (BBE) Agar Plates	Anaerobe Systems	AS-144
Laked Brucella Blood Agar w/Kanamycin and Vancomycin (LKV) Plates	Anaerobe Systems	AS-142
Brain Heart Infusion Agar (BHI)	BD Diagnostics	DF0418177
Defibrinated horse blood	Hemostat	DHB500
96-Well, Cell Culture-Treated, Flat-Bottom Microplate	Falcon	Cat#353072
Yeast Casitone Fatty Acids Broth with Carbohydrates – YCFAC Broth	Anaerobe Systems	AS-680
No MAC diet: Teklad custom diet, Glucose Only Carb (93G, Irrad)	Envigo	TD.150689
Western diet: Teklad custom diet, adjusted fat diet	Envigo	TD.96132
High MAC chow: LabDiet® JL Rat and Mouse/ Auto 6F	LabDiet	5K67
AIN-93G Basal Mix (CHO, Cellulose Free)	Envigo	TD.200788
Gelatin, bovine	Sigma	G9391

RESOURCE AVAILABILITY

Lead contact

All information and requests for further resources should be directed to and will be fulfilled by the Lead Contact, Erica Sonnenburg, erica.sonnenburg@stanford.edu.

Materials availability

This study did not generate new unique reagents.

Data and code availability

Raw data files for WGS and RNAseq can be found at Zenodo: <https://doi.org/10.5281/zenodo.7651179>. Code used to generate the figures and additional data can be found at Zenodo: <https://doi.org/10.5281/zenodo.8339517>. Isolate genomes will be available at NCBI: PRJNA1015720 upon publication. Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Bacterial culture

Bacteria not isolated in this study were purchased from DSMZ (*P. copri* DSM 18205), or ATCC (all other reference strains). Glycerol stocks were struck out on Brain Heart Infusion agar with 10% defibrinated horse blood (BHIBA) and incubated anaerobically

for 24–48 h at 37°C. All growth and culturing of *Bacteroides* and *Prevotella* strains were performed anaerobically in a Coy anaerobic chamber containing 87% N₂, 10% CO₂, and 3% H₂.

Mouse husbandry

All mouse experiments were performed in accordance with the Stanford Institutional Animal Care and Use Committee. Mice were maintained on a 12-h light/dark cycle at 20.5 °C at ambient humidity, fed *ad libitum*, and maintained in flexible film gnotobiotic isolators for the duration of all experiments (Class Biologically Clean). Swiss-Webster mice were used for gnotobiotic experiments and the sterility of germ-free mice was verified by 16S PCR amplification and anaerobic culture of feces. Sample sizes were chosen on the basis of litter numbers and controlled for sex and age within experiments. Researchers were unblinded during sample collection.⁷⁹

Statement on work with indigenous communities

In order to acquire scientific knowledge that accurately represents all human populations, rather than only reflecting and benefiting those in industrialized nations, it is necessary to involve indigenous populations in research in a legal, ethical, and non-exploitative manner.^{25,80} Here, we isolated live bacterial strains from anonymized fecal samples collected from Hadza hunter-gatherers in 2013/2014.^{4,6,81} Samples were collected with permission from the Tanzanian government, National Institute of Medical Research (MR/53I 100/83, NIMR/HQ/R.8a/Vol.IX/1542), the Tanzania Commission for Science and Technology, and with aid from Tanzanian scientists. A material transfer agreement with the National Institute for Medical Research in Tanzania specifies that collected samples are solely to be used for academic purposes. For more information on the consent practices followed, and our ongoing work to communicate the results of these projects to the Hadza, please see Carter et al.⁴ and Olm et al.⁵

METHOD DETAILS

Strain isolation from fecal samples

Samples for strain isolation were chosen from the samples reported previously based on the 16S abundance of either *Bacteroides* or *Prevotella* genera.⁶ All isolations were performed under anaerobic conditions on YCFA agar with 5% glucose or baobab powder. 1 μL of frozen feces was struck onto a single agar plate. Visible colonies from the initial plates were identified via colony PCR using bacterial 16S primers, and re-plated onto BBE and LKV plates (Anaerobe Systems). PCR products were purified using the QIAquick PCR Purification Kit (Qiagen), and sequenced via Sanger sequencing at Elim Biopharm. The resulting sequences were identified using nucleotideBLAST.⁸² Colonies that were predicted to share >95% identity with a *Bacteroides* or *Prevotella* species were re-struck two additional times on either BBE or LKV plates, respectively, to ensure a pure culture. Glycerol stocks were made by growing a liquid culture of a single colony overnight in PYG, and then mixing at a 1:1 ratio with a 50% glycerol, 50% PBS solution.

Whole genome sequencing

Genomic DNA was extracted from single-isolate cultures grown for 24 h using a MasterPure Gram Positive DNA Purification Kit. Long-read sequencing was performed using a Nanopore MinION (flow cell FLO-MIN106, Ligation Sequencing Kit SQK-LSK109; Barcoding Kit EXP-NBD104) and short read sequencing was performed using an Illumina MiSeq. Nanopore basecalling was performed with Guppy version 3.4.2m using the command “guppy_basecaller -r -i raw_fast5/-flowcell \$flowcell -kit \$kit -x auto -compress_fastq -gpu_runners_per_device 8 -q 0 -chunks_per_runner 4096”. Short read sequence quality was assessed using Fastqc with the command “fastqc -nogroup -q”, and adapters were trimmed with BBTools using the command “bbduk.sh -Xmx2g -eoom ref = adapters, phix threads = 8 ktrim = r k = 23 mink = 11 edist = 2 entropy = 0.05 tpe tbo qtrim = rl minlength = 100 trimq = 30 pigz = t unpigz = t samplerate = 0.25.” If there was more than 100x coverage of the genome, reads were normalized using the command “bbnorm.sh target = 100 min = 2”. Hybrid assembly of the short and long reads was performed using SPAdes with the command “spades.py -careful -cov-cutoff auto -k 21,33,55,77,99,127”.⁸³ RagOUT was used for chromosome-level scaffolding using either the matched reference genome of the same species for *Bacteroides* (Table 1), or *Pc H-2477* for *Prevotella*.⁵² Assembly quality was assessed with Quast.⁸⁴ Gene annotation was performed using RASTtk.⁵³

Clustering genomes into subspecies

All public *Bacteroides* and *Prevotella* genomes of “Scaffold” quality or better were downloaded from NCBI GenBank on 5/15/2023 using the program ncbi-genome-download (<https://github.com/kbclin/ncbi-genome-download>). The commands used were “ncbi-genome-download -genera *Bacteroides* -section GenBank -formats fasta -assembly-levels complete, chromosome,scaffold bacteria” and “ncbi-genome-download -genera *Prevotella* -section GenBank -formats fasta -assembly-levels complete, chromosome,scaffold bacteria” This resulted in a total of 888 and 1894 genomes of *Prevotella* and *Bacteroides*, respectively.

Public genomes were clustered along with the isolate genomes recovered in the study using dRep v3.2.1⁵⁴ using the command “dRep dereplicate -S_algorithm fastANI -sa 0.98”. The 98% ANI threshold was chosen manually based on a histogram of reported ANI values (Figure S4A). Representative genomes were chosen using dRep’s default scoring system with the following adjustments: isolates sequenced in this study were given an additional 100 points, isolate genomes used in this study were given an additional 80 points, public genomes marked as “representative genome” in Refseq were given an additional 60 points, and public genomes of “Complete Genome” and “Chromosome” quality were given an additional 40 and 20 points, respectively.

Evaluating subspecies prevalence and phylogenetic analysis

All metagenomic reads were downloaded from Carter et al.⁴ Metagenomic reads were mapped to *Prevotella* and *Bacteroides* subspecies representative genomes using Bowtie2 with default settings⁵⁵ (command “bowtie2 -x \$index -1 \$r1 -2 \$r2 | samtools sort -o \$output.bam”), and the resulting .bam files were profiled using coverM as implemented through inStrain with default settings⁵⁶ (command “inStrain profile \$bam \$fasta -s \$stb -coverm”). Genomes detected with $\geq 65\%$ genome breadth were considered “present” in a metagenome. This threshold was chosen based on manual inspection of a genome breadth histogram (Figure S4B).

The prevalence of each genome in each population was calculated as the percentage of metagenomes in which the genome was detected. Phylogenetic trees were made for *Bacteroides* and *Prevotella* subspecies representative genomes detected in at least one metagenome using GToTree v1.5.36 with the command “GToTree -H Bacteria -T IQ-TREE”. One outgroup from a different genus was included in each tree. Tree leaves were labeled based on GTDB taxonomy release r207⁵⁹. Trees were visualized using iTol.⁶⁰ Figure S1A includes GTDB representative genomes for i) all species “copri” in their species name, ii) all *Prevotella* species of isolates recovered in this study, and iii) the closest representative genome (according to GTDB) for all isolate genomes recovered in this study. For Figure S1B, 10 genomes from each clade were randomly chosen to include in the tree. *Prevotella stercorea* was included as an outgroup.

CAZyme annotation

CAZyme annotations were performed for each isolate. An additional 20 strains of *Prevotella copri* available at NCBI, with variable assembly levels, were annotated as well for comparative purpose, with the isolates and two model strains. All amino acid sequences were first compared to the full-length sequences stored in the CAZy database (Sept. 2021)⁶¹ using BlastP (version 2.3.0+).⁶² Queries obtaining 100% coverage, $>50\%$ sequence identity and E-value $\leq 10^{-6}$ were automatically annotated with the same domain composition as the closest reference homolog. All remaining sequences were subject to human curation to verify the presence of each putative module. During this process, the curator could rely on (i) bioinformatics tools, including BLAST against libraries on either full-length protein, modules only or characterized modules only, and HMMER version 3.1⁶³ against in-house built models for each CAZy (sub)family; (ii) human expertise on the appropriate coverage, sequence identity and E-value thresholds which vary across (sub)families, and ultimately on the verification of the catalytic amino acid conservation. Hierarchical clustering of isolates' CAZyme repertoires was performed using ComplexHeatmap.⁷⁵ Predicted substrate assignment was compiled from previously published works.^{6,14}

In vitro polysaccharide growth assays

Glycerol stocks were struck out on Brain Heart Infusion plates with 10% defibrinated horse blood and incubated anaerobically for 24 h at 37°C. Isolates were passaged overnight in BHI-S (*Bacteroides*), and YCFA-G (*Prevotella*). After 16h, cultures were diluted 1:50 for *Bacteroides* and 1:10 for *Prevotella* into 200 μ L of culture media in a clear, flat bottomed 96-well plate. Growth media was composed of a YCFA background, plus 0.5% carbohydrate, with the exception of inulin, which was added at a 1.5% concentration. OD600 was measured every 15 min for 48h using a BioTek Epoch2 plate reader, with 30 s of shaking prior to each reading. Normalized OD was calculated for each carbohydrate condition by subtracting the average blank OD600 from the raw OD600 for each isolate grown in the corresponding polysaccharide. Maximum OD was calculated as the highest normalized OD in the first 24h period.

Colonization and enumeration of gnotobiotic mice

For colonization with *B. thetaiotaomicron* H-2622, mice were gavaged with 300 μ L of a 3mL liquid culture grown for 16h in BHI-S. For colonization with *P. copri*, mice were gavaged with 300 μ L of a 3mL liquid culture grown for 16h in YCFAC, in which was suspended 10–15 lawns (~1 per mouse) of *P. copri* grown on BHIBA for 48 h. For *Prevotella* colonization, food was removed from mouse cages and bedding was changed 12h before gavage. Before the gavage of *Prevotella*, mice were gavaged with 300 μ L of 10% sodium bicarbonate in water. Food was returned 2h post-gavage. For bicolonization experiments, mice were first colonized with *Pc* H-2477, then gavaged with *Bt* H-2622 7 days later. Bicolonization was allowed to stabilize for 5–7 days before the diet switch.

To measure bacterial density, feces were collected from individual mice. Two biological replicates of 1 μ L feces were resuspended in 200 μ L sterile PBS, serially diluted 1:10 in sterile PBS using a 96-well tissue culture plate, and 3 technical replicates of 2 μ L of each dilution were plated on BHIBA. CFUs were counted after 36h anaerobic growth at 37 °C.

In vivo competition assays

Feces were collected from individual mice. Genomic DNA was extracted from 2 biological replicates of fecal pellets using the DNeasy PowerLyzer PowerSoil kit (Qiagen). Concentration of *Pc* and *Bt* DNA was assessed using species-specific qPCR primers (Key Resources Table). qPCR was performed using the Brilliant III, Ultra Fast SYBR Green QPCR Master Mix and a Bio Rad CFX thermocycler. Genomic DNA from *Bt* H-2622 and *Pc* H-2477 were used to generate a standard curve for each primer pair. The standard curves were used to calculate the absolute quantity of *Bt* or *Pc* DNA in the sample. The efficiency value (E) for each primer pair was calculated as $10^{(1/-\text{slope})}$ of $\log_{10}(\text{DNA input})$ against Ct value. qPCR index was calculated using this equation: E^{-Ct} primer pair.

Mouse diets

The Inulin-g and No MAC-g diets were created using 32% AIN-93G Basal Mix (CHO, Cellulose Free) and 68% carbohydrates, to match the carbohydrate content of the No MAC diet (TD.150689). The Basal Mix and carbohydrate components were suspended

in a mixture of water (1100mL per 250g package of Basal Mix) and 5% bovine gelatin as a binder. The carbohydrates (100% glucose, no MAC-g; 50% glucose and 50% inulin, Inulin-g) and gelatin were dissolved separately in MilliQ water and autoclaved. The gelatin mix and AIN-93G Basal Mix (CHO, Cellulose Free) (TD.200788) were added to the carbohydrate solution in a tissue culture hood, and the mix was allowed to solidify at 4°C. Diets are listed in the [key resources table](#). After 1 week post-colonization, standard chow was removed and replaced with the desired test diet, and the bedding was changed. Gelatin chow was replaced every 3 days as the chow dried out.

RNAseq

RNA was extracted from mouse cecal contents and *in vitro* cultures using the RNeasy PowerMicrobiome Kit (Qiagen). Ribosomal RNA depletion was performed using the RiboMinus Transcriptome Isolation Kit (Invitrogen). A cDNA library was constructed using the TruSeq Stranded Total RNA Library Prep Human/Mouse/Rat kit. Sequencing was performed on a NovaSeq SP flow cell. Quality of raw reads was assessed with Multiqc using the command “multiqc”.⁶⁴ Adapters were trimmed using Trimmomatic and the command “trimmomatic PE ILLUMINACLIP -PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36”.⁶⁵ Reads were aligned to the *Pc H-2477* and *Bt H-2622* genomes using HiSAT2 commands “hisat2-build” to generate indexes, and “hisat2 -p 8 -dta -x” to align reads to the indexes.⁶⁶ SAMtools was used to generate .bam files with the commands “samtools sort -@ 8 -o” and “samtools index”.⁶⁷ Transcripts were assembled using the Stringtie commands “stringtie”, “stringtie-merge”, and “stringtie -e -B -p 11 -G”.⁶⁸ Differential expression was analyzed using DESeq2.⁶⁹

QUANTIFICATION AND STATISTICAL ANALYSIS

Quantification and statistical analyses were performed using R version 4.2.2 or GraphPad Prism version 9 ([key resources table](#)). The ComplexHeatmap and ggplot2 packages were used to create heatmap and barplot visualizations. Prism was used to generate graphs of fecal CFU and qPCR data, to calculate the mean and standard error values for these data, and to perform statistical analyses for these experiments. The number of samples per group (n) for each experiment is indicated either in the figure legend or within the figure itself. Two-tailed Mann-Whitney tests were used to compare the distributions of two unmatched groups without the assumption of normal distribution. Paired or unpaired t-tests were used to compare two normally distributed groups of paired or unpaired samples, respectively. The Welch two-sample t test was used to compare two normally distributed groups with different standard deviations. A Fisher's exact test was used to determine the presence of nonrandom associations between two sets of categorical variables with a small sample size. The Wilcoxon test was used to determine the distinctness of the means of two groups of independent samples. Further details of statistical analyses for experiments can be found in the [results](#) section and figure legends.