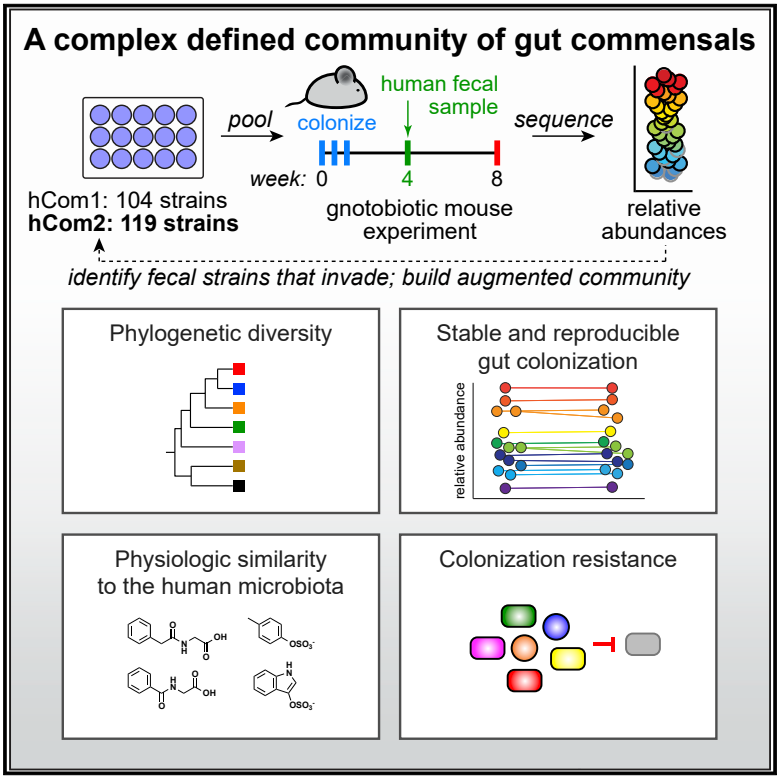
Cell

Design, construction, and *in vivo* augmentation of a complex gut microbiome

Graphical abstract



Highlights

- We introduce hCom1, a defined community of 104 gut bacterial species
- We fill open niches *in vivo* to form hCom2, a defined community of 119 species
- In gnotobiotic mice, hCom2 exhibited robust colonization resistance against *E. coli*
- Mice colonized by either hCom2 or a human fecal community are phenotypically similar

Authors

Alice G. Cheng, Po-Yi Ho, Andrés Aranda-Díaz, ..., Justin L. Sonnenburg, Kerwyn Casey Huang, Michael A. Fischbach

Correspondence

alicec2@stanford.edu (A.G.C.), kchuang@stanford.edu (K.C.H.), fischbach@fischbachgroup.org (M.A.F.)

In brief

The development of a complex community of bacteria that represents the most common taxa from the human microbiome enables further mechanistic study of genes, pathways, and species that influence host physiology and health.

Cheng et al., 2022, Cell *185*, 3617–3636
September 15, 2022
https://doi.org/10.1016/j.cell.2022.08.003

CellPress

Cell

CellPress

Resource

Design, construction, and *in vivo* augmentation of a complex gut microbiome

Alice G. Cheng,[1,8,*] Po-Yi Ho,[2,8] Andrés Aranda-Díaz,[2,8] Sunit Jain,[3] Feiqiao B. Yu,[3,4] Xiandong Meng,[3,4] Min Wang,[2,5] Mikhail Iakiviak,[2,4,5] Kazuki Nagashima,[2,4,5] Aishan Zhao,[2,4,5] Pallavi Murugkar,[4] Advait Patil,[2,4,5] Katayoon Atabakhsh,[2,4,5] Allison Weakley,[3,4] Jia Yan,[3] Ariel R. Brumbaugh,[2,4,5,9] Steven Higginbottom,[2,4,5] Alejandra Dimas,[2,4,5] Anthony L. Shiver,[2] Adam Deutschbauer,[6,7] Norma Neff,[3] Justin L. Sonnenburg,[3,5] Kerwyn Casey Huang,[2,3,4,5,*] and Michael A. Fischbach[2,3,4,5,10,*]

[1]Department of Gastroenterology & Hepatology, Stanford School of Medicine, Stanford, CA 94305, USA
[2]Department of Bioengineering, Stanford University, Stanford, CA 94305, USA
[3]Chan Zuckerberg Biohub, San Francisco, CA 94158, USA
[4]ChEM-H Institute, Stanford University, Stanford, CA 94305, USA
[5]Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford University, Stanford, CA 94305, USA
[6]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[7]Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA 94720, USA
[8]These authors contributed equally
[9]Present address: Gladstone Institutes Office of Corporate Liaisons and Ventures, 650 Owens Street, San Francisco, CA 94158
[10]Lead contact
*Correspondence: alicec2@stanford.edu (A.G.C.), kchuang@stanford.edu (K.C.H.), fischbach@fischbachgroup.org (M.A.F.)
https://doi.org/10.1016/j.cell.2022.08.003

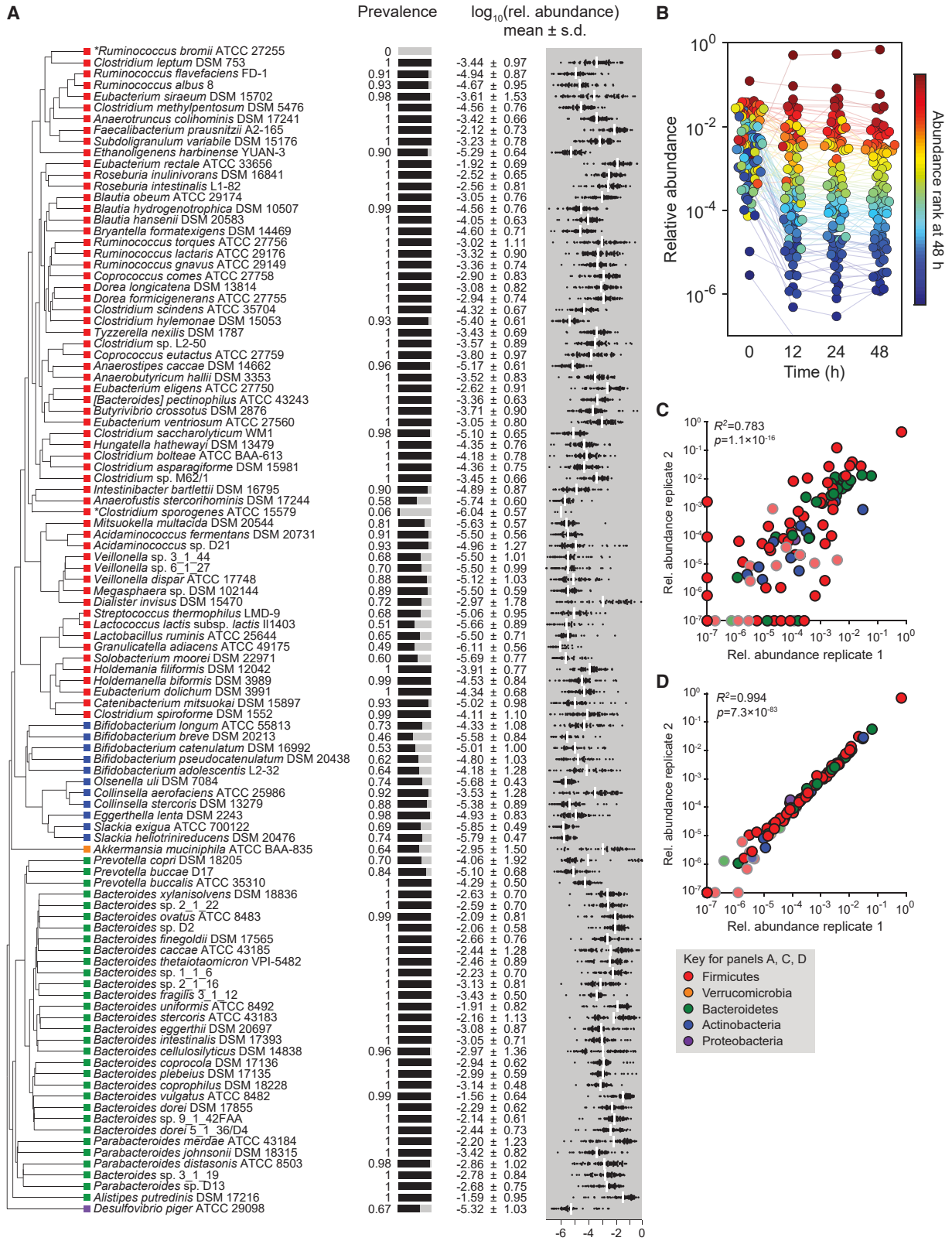
SUMMARY

Efforts to model the human gut microbiome in mice have led to important insights into the mechanisms of host-microbe interactions. However, the model communities studied to date have been defined or complex, but not both, limiting their utility. Here, we construct and characterize *in vitro* a defined community of 104 bacterial species composed of the most common taxa from the human gut microbiota (hCom1). We then used an iterative experimental process to fill open niches: germ-free mice were colonized with hCom1 and then challenged with a human fecal sample. We identified new species that engrafted following fecal challenge and added them to hCom1, yielding hCom2. In gnotobiotic mice, hCom2 exhibited increased stability to fecal challenge and robust colonization resistance against pathogenic *Escherichia coli*. Mice colonized by either hCom2 or a human fecal community are phenotypically similar, suggesting that this consortium will enable a mechanistic interrogation of species and genes on microbiome-associated phenotypes.

INTRODUCTION

Experiments in which a microbial community is transplanted into germ-free mice have opened the door to studies of mechanism and causality in the microbiome. These efforts fall into two categories based on the nature of the transplanted community: complete, undefined communities (i.e., fecal samples) versus incomplete but defined communities (i.e., synthetic communities). Fecal transplantation studies have shown that the microbiome plays a role in a variety of host phenotypes including the response to cancer immunotherapy (Gopalakrishnan et al., 2018; Matson et al., 2018; Routy et al., 2018), caloric harvest (Ridaura et al., 2013), colonization resistance to enteric pathogens (Buffie et al., 2015), and neural development (Buffington et al., 2021; Sharon et al., 2019). Although illuminating, a limitation of this format is that it is difficult to "fractionate" an undefined community, making it challenging to discover which species are involved in a phenotype of interest.

Synthetic communities are less well developed as model systems for the gut microbiome (Blasche et al., 2017; Pacheco and Segrè, 2019; Walter et al., 2018; Widder et al., 2016; Xavier, 2011). Pioneering efforts have shown that a synthetic community can model the impact of diet on the microbiome (Faith et al., 2011), identified genes required for *Bacteroides thetaiotaomicron* growth in the mouse intestine in the presence of a 15-member community (Goodman et al., 2009), and demonstrated that complex communities composed of species isolated from a single donor can stably colonize mice (Goodman et al., 2011). More recent studies with defined communities have revealed mechanistic insights into immune modulation, glycan consumption, and other complex phenotypes driven by the microbiome (Faith et al., 2014; van der Lelie et al., 2021; Patnode et al., 2019; Wymore Brand et al., 2015). Although synthetic communities enable precise control over composition and manipulations such as strain dropouts and gene knockouts, the communities used are typically of low complexity (<20 strains),



(legend on next page)

limiting their ability to model the biology of a native-scale microbiome.

An ideal model system for the gut microbiome would capture the advantages of both approaches: near-native complexity would allow a model microbiome to capture properties of an ecosystem that are missing from simpler model systems, including emergent phenomena such as resilience to perturbation (Dethlefsen and Relman, 2011; Ng et al., 2019) and cooperative metabolism (Morris et al., 2013). Moreover, complex consortia are a promising starting point for *in vivo* studies of the gut microbiome, for which they are better suited to model community-level phenomena such as immune modulation and the formation of structured multispecies biofilms.

Complete definition (i.e., communities composed entirely of known organisms) would enable reductionist experiments to probe mechanism. The ability to construct communities with defined composition is especially relevant in the context of experiments testing whether phenotypes can be transferred to germ-free mice via fecal transplant (Gopalakrishnan et al., 2018; Ridaura et al., 2013; Routy et al., 2018). At present, since transplanted communities are typically undefined, it is difficult to uncover the mechanisms underlying these phenomena. A defined model system of sufficient complexity would enable reductionist follow-up experiments, bringing the gut microbiome in line with other model systems in which mechanistic studies are possible.

To this end, we sought to create a community that is defined, enabling precise manipulations, and complex enough to exhibit emergent features of a complete community such as stability upon engraftment and colonization resistance. We started by constructing a complex defined community that contains the most prevalent bacterial species in the human gut microbiome (hCom1). We demonstrate that the assembly of this 104-member community is reproducible even for very low-abundance species. By systematically perturbing this community and its growth medium, we uncover strain-nutrient and strain-strain (e.g., syntrophic) interactions that underlie its composition. We then colonize germ-free mice with hCom1, showing that it adopts a stable, highly reproducible configuration in which its constituent species span six orders of magnitude of relative abundance. We augment the community by filling open niches using an iterative, ecology-based process and show that the expanded community (hCom2) is more resilient to perturbation and resistant to pathogen colonization. Finally, we demonstrate that mice colonized by hCom2 are phenotypically similar to mice harboring an undefined human fecal sample, suggesting that our consortium and augmentation process lay the foundation for developing complete, defined models of the human gut microbiome.

RESULTS

Designing and building a complex synthetic community

We set out to design a community composed of the most common bacterial species in the human gut microbiome. We analyzed metagenomic sequence data from the NIH Human Microbiome Project (HMP) to determine the most prevalent organisms—those that were present in the largest proportion of subjects, regardless of abundance. Although the HMP is not broadly representative of microbiomes from diverse geographies and ethnicities (Deschasaux et al., 2018; He et al., 2018; Sonnenburg and Sonnenburg, 2019), this data set was well suited to our purposes since it was sequenced at very high depth, enabling us to identify low-abundance organisms that are nevertheless highly prevalent (Kraal et al., 2014). After rank-ordering bacterial strains by prevalence, we found that ~20% (166/844) were present in >45% of the HMP subjects. Of these 166 strains, we were able to obtain 99 from culture collections or individual laboratories (Figure 1A; omitted strains are listed in Table S1). The profiled strains of three additional species were unavailable; instead, we used alternative strains of the same species (*Lactococcus lactis* subsp. *lactis* Il1403, *Bacteroides xylanisolvens* DSM 18836, and *Megasphaera* sp. DSM 102144). We introduced two additional strains to enable downstream experiments: *Ruminococcus bromii* ATCC 27255, a keystone species in polysaccharide utilization (Ze et al., 2012); and *Clostridium sporogenes* ATCC 15579, a model gut *Clostridium* species for which genetic tools are available (Dodd et al., 2017; Funabashi et al., 2020; Guo et al., 2019). These 104 strains—a community termed "hCom1"—are prevalent and abundant in Western human gut communities (Data S1). Notably, unlike other defined communities used to model the gut microbiome, our consortium is within ~2-fold of the estimated number of species in a typical human gut (STAR Methods) (Faith et al., 2013; Qin et al., 2010).

Figure 1. A complex gut bacterial community

(A) A phylogenetic tree of the 104 strains in hCom1 based on a multiple sequence alignment of conserved single-copy genes. The community was designed by identifying the most prevalent strains in sequencing data from the NIH Human Microbiome Project (HMP). Colored squares indicate the phylum of each strain: Firmicutes, red; Actinobacteria, blue; Verrucomicrobia, orange; Bacteroidetes, green; and Proteobacteria, purple. Also shown are the prevalence and relative abundances of each strain in the data set from the NIH HMP (n = 81 subjects). The prevalence is the fraction of subjects in which the strain was detected. The distribution of $\log_{10}$ (relative abundance) across subjects is shown with the mean denoted by a white line for each strain. *Ruminococcus bromii* ATCC 27255 and *Clostridium sporogenes* ATCC 15579 were added to the community despite low prevalence in the HMP samples.

(B) The community reaches a stable configuration quickly. The community was propagated *in vitro* in SAAC medium to test the stability of its composition. Each dot is an individual strain; the collection of dots in a column represents the community at a single time point. Strains are colored according to their rank-order abundance in the community at 48 h. By 12 h, the relative abundances of strains in the community spanned six orders of magnitude and remained largely stable through 48 h.

(C) Communities generated from two inocula prepared on different days (i.e., biological replicates) have a similar architecture at 48 h.

(D) Communities generated from the same inoculum (i.e., technical replicates) have a nearly identical composition at 48 h. In (C) and (D), the color of each circle represents the phylum of the corresponding species, and circles with gray outlines and faint colors represent strains whose presence could be explained by read mis-mapping.

See also Tables S1 and S2 and Data S1.

A streamlined strain growth protocol simplified the assembly of hCom1 and single-strain dropouts (STAR Methods). We found that each of our 104 strains can be propagated in Mega Medium (MM), Chopped Meat Medium (CMM), or both (key resources table). Growth rates, carrying capacities, and time of entry into stationary phase varied widely across strains and media. To simplify the process of community assembly while ensuring that slow-growing strains were actively dividing, each strain was inoculated from a frozen stock into liquid medium and passaged every 24 h for a total of 2–3 days. Before mixing individually cultured strains, we adjusted the volumes of each culture to achieve similar optical densities. A subset of the strains did not reach the diluted culture density of the remaining strains (STAR Methods); we added these cultures undiluted. We confirmed that our starting cultures were pure using metagenomic sequencing and high accuracy read mapping, as described in the next section.

Development of a highly accurate metagenomic read-mapping pipeline

Having assembled a community of 104 species, we next addressed how to quantify the abundance of each strain accurately, a major challenge given our expectation that some strains would be present at low abundance. Various strains in the community have identical 16S hypervariable sequences in the V3-V4 region, ruling out 16S amplicon-based methods. We considered designing a custom amplicon-based pipeline, but such an approach would require the design and validation of new primer sets for future communities. As an alternative, we sought to use metagenomic sequencing to quantity community composition.

To test the performance of existing metagenomic analysis tools, we generated three "ground truth" data sets. The first two consisted of simulated reads generated from the assembled genome sequences of each strain: one in which all 104 strains were equally abundant (to test sensitivity and specificity), and another in which strain abundance varied over six orders of magnitude (to test dynamic range). The third set consisted of actual reads derived from sequencing each strain individually using the same protocol as in subsequent community analyses. This data set allowed us to account for biases introduced by library construction and sequencing.

We found that metagenomic read mappers based on a combination of Bowtie2 (Langmead and Salzberg, 2012) and SAMtools (Li et al., 2009) were sensitive but inaccurate: there was substantial mis-mapping of reads from one strain to others, such that whole-genome sequencing data from an individual strain was often interpreted as having arisen from multiple strains. Read mis-mapping from any abundant strain could therefore create noise that exceeds signal from low-abundance strains, degrading accuracy. In contrast, algorithms that focus on a few universal genes or unique k-mers such as MetaPhlAn2 (Truong et al., 2015), MIDAS (Nayfach et al., 2016), Kraken2/Bracken (Lu et al., 2017; Wood et al., 2019), IGGsearch (Nayfach et al., 2019), or Sourmash (Titus Brown and Irber, 2016) were generally accurate to the species level, but since they only use a small fraction of the reads (<1%), their ability to detect low-abundance or closely related strains is limited.

To address these challenges, we developed a new algorithm, NinjaMap (Data S2). Taking advantage of the fact that every strain in our community has been sequenced (Table S2), NinjaMap can quantify strain abundances with high accuracy across six orders of magnitude (STAR Methods). In brief, NinjaMap considers every read from a sample. If a read does not match perfectly to any of the genomes in the community (typically 3%–4% of the reads), it is tabulated but not assigned. If a read has a perfect match to only one strain, it is assigned unambiguously to that strain. If a read matches more than one strain perfectly, it is temporarily placed in escrow. After all unambiguous assignments are made, an initial estimate of the relative abundance of each strain is computed. Reads in escrow are then fractionally assigned in proportion to the relative abundance of each strain, normalized by the total size of the genomic regions available for unique mapping to avoid bias in favor of strains with large or phylogenetically distinct genome sequences. Finally, relative abundances are computed.

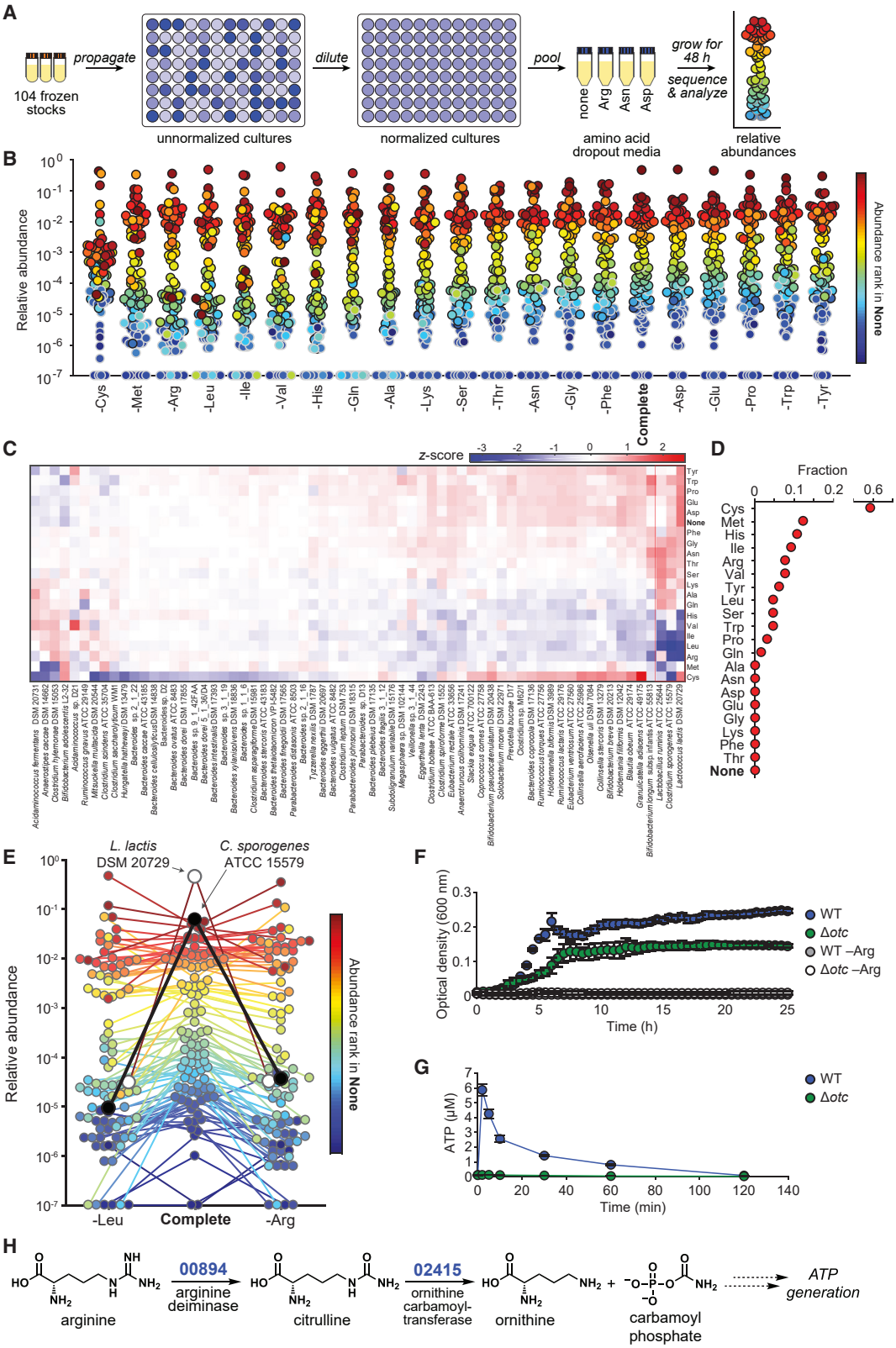
To assess the performance of NinjaMap, we conducted two tests. First, we assessed the degree of read mis-mapping from and into each strain's ledger. We quantified how many reads from strain 1 were mis-assigned to strains 2–104 (which would underestimate the abundance of strain 1 in a community), and how many reads from strains 2–104 were mis-assigned to strain 1 (which would overestimate the abundance of strain 1). For simulated reads, most instances of these two types of read mis-mapping collectively resulted in relative abundance errors $\lesssim 10^{-5}$ (Data S2; STAR Methods). For actual reads, mis-mapping was more frequent but still typically below a threshold of 10^{-4} (i.e., 0.01% relative abundance); mis-mapping likely arose either from deviations between the database genome sequence and the actual sequence of the strain in our collection, or from the process of sample preparation and sequencing (Data S2; STAR Methods). The expected contribution to relative abundance from mis-mapping in a community context can be even lower for some strains (Data S2).

Second, we used NinjaMap to analyze simulated reads from a 104-strain community. We found that this tool can accurately quantify strains with abundances as low as 10^{-6} in the context of a mixed community of known composition (Data S2), in agreement with the analysis of single-isolate samples. Thus, NinjaMap is capable of quantifying strains accurately over a wide dynamic range of relative abundances.

Community construction is highly reproducible

We began by measuring the degree of reproducibility in community composition data by constructing and propagating the 104-member community multiple times *in vitro*. We included technical replicates to assess variation in bacterial growth, DNA extraction, and sequencing and biological replicates to determine the impact of differences in the preparation of the inocula. We propagated the communities for 48 h and extracted DNA for sequencing at 0, 12, 24, and 48 h.

The range of cell densities at $t = 0$ spanned multiple orders of magnitude (Figure 1B), with a mean $\log_{10}$(relative abundance) of -2.5 ± 0.8 for all detectable strains. 95/104 strains were detectable at $t = 0$; the remaining strains, which grew poorly when cultured individually, were below the limit of detection or had



(legend on next page)

abundances that could potentially be explained by read mis-mapping. The communities reached a relatively stable configuration by 12 h (Figure 1B), with a remarkable degree of reproducibility among biological replicates (Figure 1C). Notably, very low-abundance strains ($<10^{-4}$) were only slightly more variable than high-abundance strains. Technical replicates were even more similar (Figure 1D), indicating that community growth, DNA extraction, and sequencing contributed only modestly to variability. Taken together, these results indicate that community composition is robust to experimental variation.

A nutrient dropout screen to map strain-nutrient interactions in the community

We next sought to explore the network of strain-nutrient interactions in the community. Although much is known about polysaccharide foraging by gut commensals (Martens et al., 2014), far less is known about amino acid utilization; hence, we performed the experiment in a defined growth medium (standard amino acid complete [SAAC]; STAR Methods) from which we could remove one amino acid at a time. Since amino acids are often utilized in pairs (Nisman, 1954; Smith and Macfarlane, 1997), eliminating one at a time from a complete background rather than adding one at a time to a null background has greater potential to reveal phenotypes relevant to community function. Moreover, performing this screen in the context of a diverse community (as opposed to the traditional practice of analyzing the growth of isolated strains) enables the potential study of community-dependent effects such as nutrient competition or mutualism-dependent nutrient utilization.

To map strain-amino acid interactions, we constructed the 104-member community (STAR Methods) and used it to inoculate 20 defined growth media, each deficient in a single amino acid, as well as complete SAAC (Figure 2A). Samples were taken at 48 h and metagenomic sequencing data were analyzed to determine the impact of amino acid deficiency on the relative abundance of each strain.

Global analysis of strain-amino acid interactions

To identify strain-amino acid interactions, we tabulated strains whose relative abundance deviated significantly from the mean across conditions, taking advantage of the fact that most amino acid dropouts had little effect on most strains (Figure 2B; STAR Methods). When the community was propagated in the complete defined medium, relative abundances spanned >6 orders of magnitude. 36% of the strains were present at 10^{-4} to 10^{-2} relative abundance, 8 strains were $>10^{-2}$, and 50 were $<10^{-4}$ (Figure 2B). In agreement with simulated results, NinjaMap was sensitive to strains with relative abundances as low as 10^{-6}, enabling us to quantify the 56% of strains that were below the 10^{-3} limit of detection commonly used for metagenomic analyses (Franzosa et al., 2015). Our system is therefore capable of studying low-abundance microbes, some of which are known to have large biological impacts (Buffie et al., 2015; Funabashi et al., 2020).

To identify significant responses, we calculated the standard deviation of the relative abundance of each strain across experiments and computed Z scores (Figure 2C; STAR Methods). Strain-amino acid interactions that were previously identified in monoculture studies were also observed in our community format. *Anaerostipes caccae*, whose growth is stimulated by methionine (Soto-Martin et al., 2020), decreased in relative abundance in a community grown in methionine-deficient medium ($Z = -3.48$). Likewise, *C. sporogenes* expansion was impeded by the absence of leucine ($Z = -2.56$), a substrate it oxidatively decarboxylates to isovalerate to generate electrons (Guo et al., 2019). These observations demonstrate that although >100 strains are competing for the same nutrients, the effects of eliminating one amino acid on the growth of one strain are readily observable in the context of a complex and diverse community.

Most strains responded to amino acid removal in ≤ 4 cases (Figure 2B). Moreover, relative abundances displayed low variability, with a mean standard deviation of $\log_{10}$(relative abundance) across strains <0.43. Only three strains, all of which are

Figure 2. Systematic analysis of strain-amino acid interactions

(A) Schematic of the amino acid dropout experiment. Frozen stocks of the 104 strains were used to inoculate cultures that were grown for 24 h, diluted to similar optical densities (to the extent possible), and pooled. The mixed culture was used to inoculate one of twenty defined media lacking one amino acid at a time. After 48 h, communities were sequenced and analyzed by NinjaMap to determine changes relative to growth in the complete defined medium.

(B) Community composition is impacted by amino acid dropout. Each dot is an individual strain; the collection of dots in a column represents the community at a single time point. Strains are colored according to their rank-order abundance in the community grown in complete defined medium (SAAC). Strains whose relative abundance could be explained by read mis-mapping from a more abundant strain in the same sample are plotted with a gray outline. Undetected strains were set to 10^{-7} for visualization.

(C) Heat map showing the hierarchically clustered Z scores for each strain (x axis) across amino acid dropouts (y axis). The Z score was calculated based on the standard deviation of strain abundance across all samples except the cysteine dropout (STAR Methods). The Firmicutes *L. lactis*, *C. sporogenes*, and *L. ruminis* grew less robustly in the absence of Leu and Ile. Strains whose abundances could be explained by mis-mapping from a higher-abundance strain are not shown.

(D) The effect of amino acid removal varies widely across amino acids. The fraction of strains with $|Z| > 2$ is shown for each amino acid dropout (n = 66).

(E) The absence of leucine or arginine leads to a large decrease in *C. sporogenes* relative abundance. Strains are colored according to their rank-order abundance in the community grown in complete defined medium. Only strains that were detected in at least one of the three samples were included (n = 92). *C. sporogenes* is highlighted in black. *L. lactis* is highlighted in white. Undetected strains were set to 10^{-7} for visualization.

(F) *C. sporogenes* growth in complete defined medium is dependent on the presence of arginine (Arg), and ornithine transcarbamoylase (*otc*) is partially responsible for Arg metabolism. Wild-type *C. sporogenes* and a Δ*otc* mutant were grown in complete defined medium ± Arg. Growth curves depict the mean of 3 replicates. Error bars represent 1 standard deviation.

(G) *C. sporogenes* requires *otc* to produce ATP from arginine. Intracellular ATP levels in *C. sporogenes* incubated in PBS containing 2 mM Arg are shown.

(H) A proposed pathway for Arg metabolism in *C. sporogenes*. Based on these data, we propose that Arg is converted to citrulline by the putative Arg deiminase CLOSPO_00894; citrulline is then hydrolyzed to ornithine and carbamoyl phosphate by the putative ornithine transcarbamoylase CLOSPO_02415, leading to the production of ATP.

See also Tables S3, S5, and S6 and Data S3.

Firmicutes, were responsive to removal in >4 cases: *Lactococcus lactis* DSM 20729, *Clostridium sporogenes* ATCC 15579, and *Lactobacillus ruminis* ATCC 25644 (Data S3; Table S3). Thus, under these growth conditions, most strains are largely insensitive to amino acid removal, whereas a small minority are highly responsive. We note that the response of a strain to amino acid removal may be direct (e.g., due to utilization for energy) or indirect (e.g., amino acid removal impacts an interacting strain).

Amino acids varied widely in terms of their impact on community composition (Figure 2D). More than half of the strains responded to cysteine removal, likely due to its effect as a reducing agent. More than 5% of the strains responded to methionine, histidine, isoleucine, arginine, valine, and tyrosine removal, whereas for eight amino acids, there were no significant changes to the community at all (Figure 2D). Interestingly, there were large differences among similar amino acids: no strains responded to lysine removal, whereas 10.6% and 7.6% of the strains responded to histidine and arginine removal, respectively. The removal of isoleucine, leucine, and arginine had a particularly large impact on community structure: *C. sporogenes* and *L. lactis*, the two most abundant strains when grown in complete defined medium, decreased >500-fold in relative abundance when any of these amino acids were removed (Figure 2E); this sensitivity was also observed in a biological replicate experiment (Data S3). Taken together, our data suggest that certain amino acids are "keystone" nutrients that play an important role in determining community composition.

C. sporogenes uses arginine to generate ATP

Among the 86 candidate strain-amino acid interactions revealed by our screen, we were particularly intrigued by those involving *C. sporogenes*. Although *C. sporogenes* can oxidize and reduce aromatic amino acids (Dodd et al., 2017), its relative abundance was unaffected by the removal of phenylalanine, tyrosine, or tryptophan (Data S3). In contrast, the removal of leucine, isoleucine, and arginine each had large impact on the fitness of *C. sporogenes* in the community. The second strongest phenotype was a decrease in relative abundance in the absence of arginine (Figures 2E and Data S3); although *C. sporogenes* is known to metabolize arginine (Venugopal and Nadkarni, 1977; Wildenauer and Winter, 1986), no impact of arginine on growth or energy metabolism had been observed in prior work. To validate and characterize this interaction, we compared *C. sporogenes* growth in complete defined versus arginine-deficient medium. Although *C. sporogenes* grew well in complete defined medium, it exhibited a large growth defect in the absence of arginine (Figure 2F), indicating that this amino acid is an important substrate for growth.

C. sporogenes can use other amino acids as substrates to support ATP synthesis (Dodd et al., 2017). Hypothesizing that the same is true for arginine, we incubated wild-type *C. sporogenes* in a culture medium deficient in substrates for ATP synthesis. Upon addition of arginine, intracellular ATP levels rose sharply (Figure 2G), indicating that *C. sporogenes* generates ATP (directly or indirectly) from arginine.

To identify the enzymes involved in this process, we parsed the *C. sporogenes* genome for pathways known to capture energy from arginine. This search yielded candidate genes for each of the three steps in the arginine deiminase pathway (Figure 2H), which catalyzes the net conversion of arginine to ornithine plus CO_2 and two equivalents of ammonium, generating one equivalent of ATP (Cunin et al., 1986). Using a method we recently developed to construct scarless deletions in *C. sporogenes* (Guo et al., 2019), we generated strains deficient in the putative arginine deiminase (CLOSPO_00894, Δ*adi*) or ornithine carbamoyltransferase (CLOSPO_02415, Δ*otc*). The Δ*otc* mutant was unable to generate ATP in response to arginine provision, consistent with a role for the arginine deiminase pathway in *C. sporogenes* energy production (Figure 2G). In contrast, the Δ*adi* mutant showed no defect in arginine-induced ATP production (Data S3), suggesting the possibility of an alternative pathway to generate citrulline from arginine. Consistent with these observations, the Δ*otc* mutant (but not the Δ*adi* mutant) was growth-deficient in complete defined medium (Figure 2F; Data S3). The deficiency was partial, suggesting that an alternative pathway can generate energy from arginine under these conditions. Together, these results show that arginine metabolism by the arginine deiminase pathway contributes directly to the cellular ATP pool, augmenting our understanding of how amino acid metabolic pathways contribute to the fitness of a gut commensal within a complex community.

Attributes of a complex defined community in gnotobiotic mice

Our central goal in designing hCom1 was to enable mechanistic studies of the microbiome in the context of host colonization. As a starting point for *in vivo* work, we colonized germ-free Swiss-Webster (SW) mice with hCom1 (Figure 3A), which we prepared by propagating each strain individually and mixing OD-normalized cultures (STAR Methods). We sampled fecal pellets from the mice weekly for 8 weeks, enumerated community composition in the inoculum and each fecal sample by metagenomic sequencing, and performed read analysis using NinjaMap.

Our analysis yielded two main conclusions. First, almost all strains in the inoculum colonized the mouse gut (Figures 3B and 3C). We confirmed the presence of 103/104 strains in the inoculum; of these, 101 strains were detected in the mice at least once. The three strains we failed to detect in mice—*Ethanoligenens harbinense* YUAN-3, *Clostridium methylpentosum* DSM 5476, and *Ruminococcus albus* 8—were slow growing and difficult to cultivate. Although strain relative abundances spanned >6 orders of magnitude, nearly all strains exhibited low variation across 20 mice in four cages, with coefficient of variation (CV, standard deviation/mean) <0.4.

Second, the community quickly reached a stable configuration (Figure 3D). Averaged across mice, relative abundances remained largely constant 2 weeks after colonization, with Pearson's correlation coefficient >0.95 at each time point with respect to the composition in week 8. After the first week, relative abundances stayed within a narrow range for the duration of the experiment (mean CV<0.2 across the 96 strains that remained above the limit of detection). Large shifts in relative abundance were rare: only 27/312 (8.7%) week-to-week strain-level changes were >10-fold.

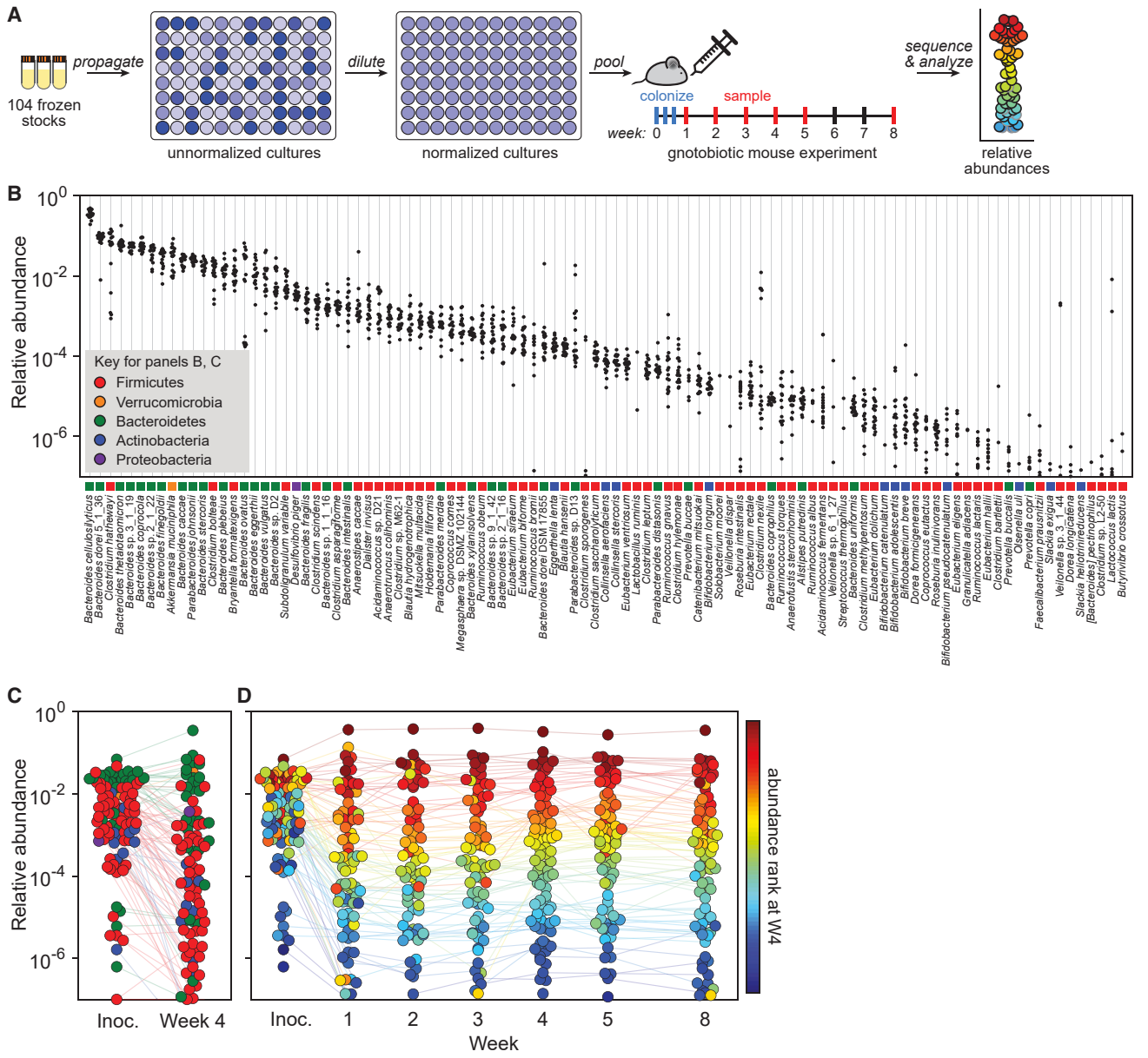


Figure 3. Colonizing germ-free mice with a complex gut bacterial community

(A) Schematic of the experiment. Frozen stocks of the 104 strains were used to inoculate cultures that were grown for 24 h, diluted to similar optical densities (to the extent possible; STAR Methods), and pooled. The mixed culture was used to colonize germ-free Swiss-Webster (SW) mice by oral gavage. Fecal samples were collected weekly at weeks 1–5 and week 8, subjected to metagenomic sequencing, and analyzed by NinjaMap to measure the composition of the community at each time point.

(B) Relative abundances for most strains are tightly distributed. Each column depicts the relative abundance of an individual strain across all mice at week 4.

(C) Average relative abundances of the inoculum versus the communities at week 4. Strains in the community spanned >6 orders of magnitude of relative abundance when colonizing the mouse gut. Dots are colored by phylum according to the legend in (B). Data represent the average of all mice in the experiment.

(D) hCom1 reaches a stable configuration by week 2. Each dot is an individual strain; the collection of dots in a column represents the community at a single time point averaged over 5 mice co-housed in a cage. Strains are colored according to their rank-order relative abundance at week 4.

An ecology-based process to fill open niches in the community

Although hCom1 is composed of prevalent species from the human gut microbiome, it is not as complex or phylogenetically rich as a human fecal community; the process that dictated its membership was not designed to ensure completeness by any functional or ecological criteria. To create a defined community that better models the gut microbiome, we sought to augment hCom1 by increasing the number of niches it fills in the gastrointestinal tract (Figure 4A). We designed an experimental strategy

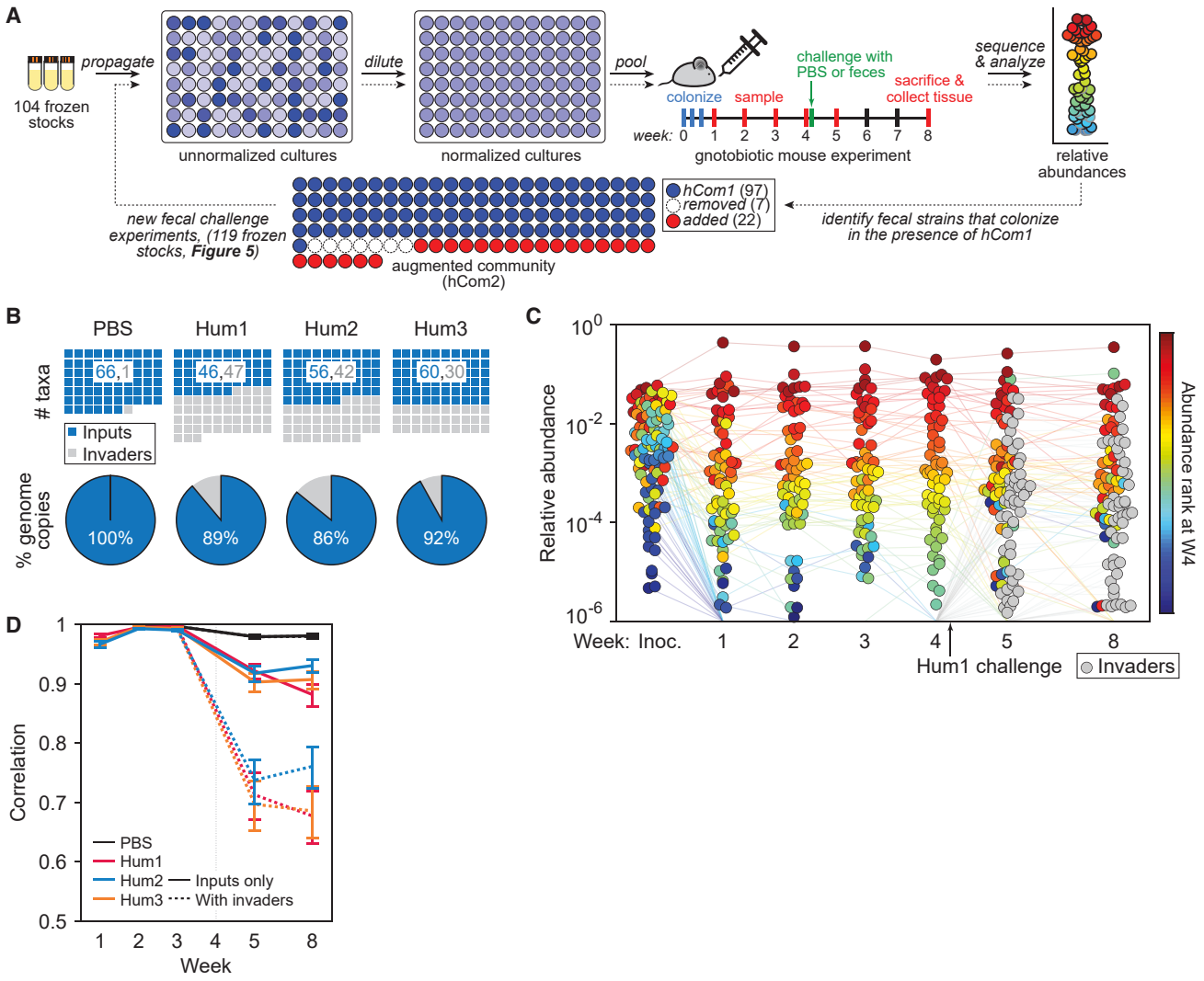


Figure 4. Challenging hCom1 with human fecal communities to identify strains that fill open niches

(A) Schematic of the experiment. Mice were colonized by freshly prepared hCom1 and housed for 4 weeks, presumably filling the metabolic and anatomical niches accessible to the strains in the community. At the beginning of week 5, the mice were challenged with one of three fecal communities from a healthy human donor or with PBS as a control; we reasoned that fecal strains that would otherwise occupy a niche already filled by hCom1 would be excluded, whereas fecal strains whose niche was unfilled would be able to cohabit with hCom1. After 4 additional weeks, we used metagenomic sequencing coupled with MIDAS to analyze community composition from fecal pellets collected at weeks 1–5 and 8. We then identified strains that colonized in the presence of hCom1 to augment the community to create hCom2, which was then used for another round of challenge experiments (Figure 5).

(B) hCom1 is broadly but not completely resistant to fecal challenge. All plots represent MIDAS bins, a rough proxy for species-level taxa. Top row: blue squares in the waffle plots indicate species that derive from hCom1, and gray squares represent species from the fecal communities. Bottom row: pie charts representing the total relative abundance of MIDAS bins that derive from hCom1 versus the fecal communities. An average of 89% of the genome copies from week 8, comprising 58% of the MIDAS bins, derived from hCom1. The remaining 11% of the genome copies, and 42% of the MIDAS bins, represent new species that joined hCom1 from one of the fecal samples.

(C) Despite the addition of new strains, the architecture of the community remains intact. Each dot is an individual strain; the collection of dots in a column represents the community at a single time point averaged over the 5 co-housed mice that were challenged with fecal community Hum1. Strains are colored according to their rank-order relative abundance at week 4. Gray circles represent invading species derived from fecal community Hum1, defined as any species not present in weeks 1–4 in the group of mice shown.

(D) The relative abundances of the hCom1-derived species present post-challenge are highly correlated with their pre-challenge levels. Pearson's correlation coefficient with respect to the average relative abundance in weeks 2 and 3 are shown for the PBS control and 3 fecal community challenges, averaged across mice that received the same challenge. Correlation coefficients are shown for the 104 hCom1 species (solid lines) and for all species including invaders (dashed lines).

See also Figures S1, S2, and S3, Tables S4 and S7, and Data S4.

based on the principle of colonization resistance (Buffie and Pamer, 2013; Lawley and Walker, 2013), an ecological phenomenon in which resident organisms exclude invading species from occupied niches. We colonized germ-free mice for 4 weeks with hCom1, presumably filling the metabolic and anatomical niches in which its species reside. We then challenged these mice with one of three undefined fecal samples (Hum1–3), reasoning that invading species that would otherwise occupy a niche already filled by hCom1 would be excluded, whereas invading species whose niche was unfilled would be able to cohabit with hCom1. After four additional weeks, we used metagenomic sequencing to analyze community composition from fecal pellets.

To determine which species from each fecal sample colonized in the presence of hCom1, we analyzed the composition of fecal pellets collected in weeks 5–8 to assign species as "input" (hCom1-derived) or "invader" (fecal sample-derived). For this analysis, we used MIDAS (Nayfach et al., 2016), an enumeration tool that—unlike NinjaMap—does not require prior knowledge of the constituent strains. MIDAS and NinjaMap reported highly concordant relative abundance profiles using sequencing reads from hCom1-colonized mice, although—as expected—MIDAS was less sensitive since it utilizes only 1% of sequencing reads (STAR Methods; Data S4). We used MIDAS for subsequent analyses of samples that were partially or completely undefined.

Using MIDAS, we cannot determine whether a strain present both pre- and post-challenge was derived from hCom1 (i.e., the original strain colonized persistently) or the fecal sample (i.e., a new strain displaced the original strain). To gain further insight into strain displacement versus persistence, we recruited reads from samples taken 4 weeks post-challenge (week 8) to a database composed of the hCom1 genome sequences, using only reads that were 100% identical to one or more of the genomes. We focused our analysis on genomes with high depth of coverage ($\geq$10X). More than 60% of these strains were covered broadly ($\geq$95%) by perfectly matching reads, indicating that most strains present pre- and post-challenge were either hCom1-derived or a closely related strain (Data S4).

As expected, mice challenged by saline instead of a fecal sample showed no evidence of new species post-challenge (Figure 4B). In hCom1-colonized mice challenged by a fecal sample, an average of 89% of the genome copies from week 8 (and 58% of the MIDAS bins, a rough proxy for species) derived from hCom1 (Figure 4B). The remaining 11% of the genome copies (and 42% of the MIDAS bins) represent new species that joined hCom1 from one of the fecal samples. Despite the addition of new species, the architecture of the community remained intact (Figure 4C): the relative abundances of the hCom1-derived species present post-challenge were highly correlated with their pre-challenge levels (Pearson's r >0.85) (Figure 4D). Thus, hCom1 is broadly but not completely resilient to a human fecal challenge.

Designing and constructing an augmented community

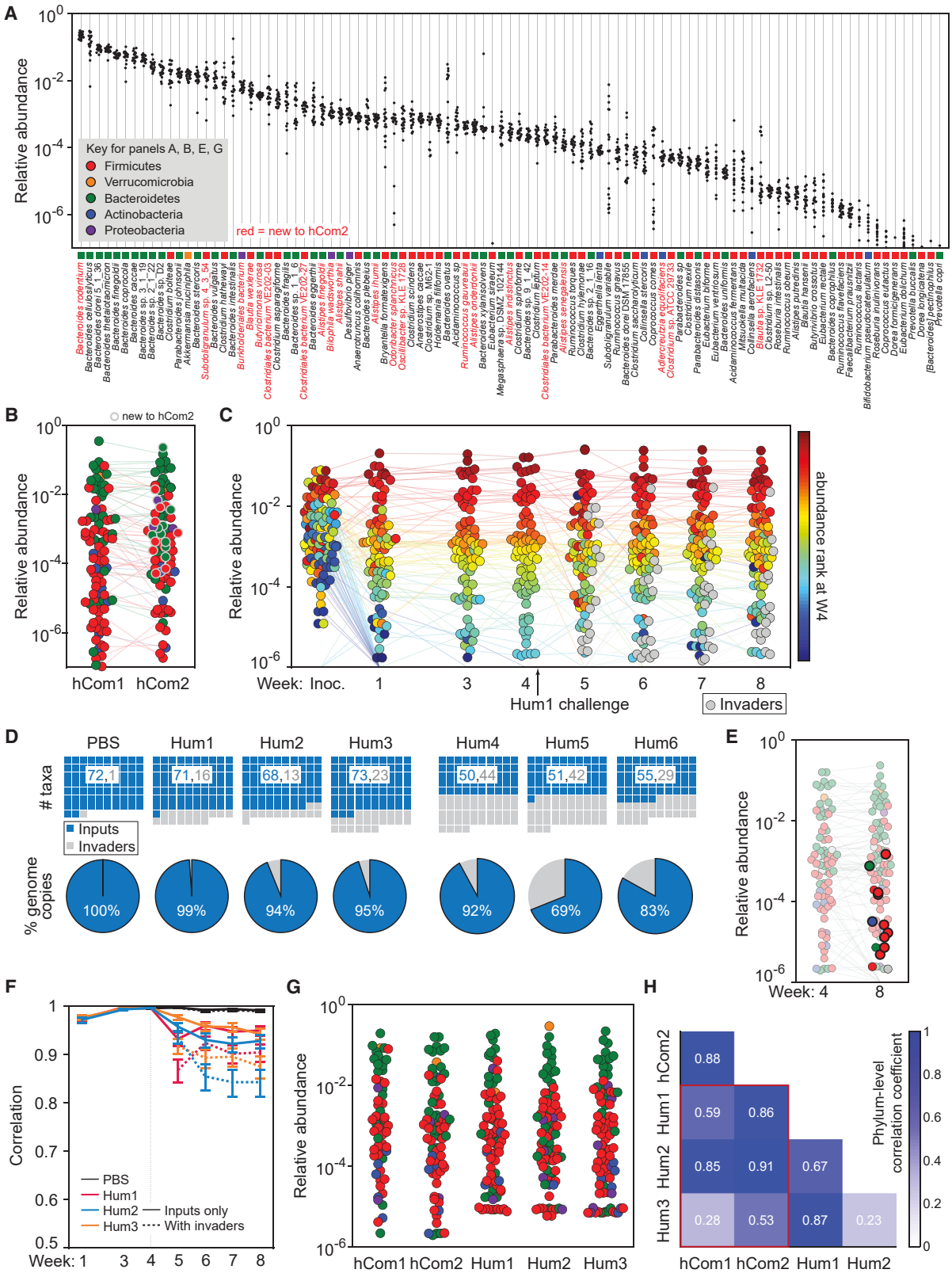
The observation that only a small fraction of the post-challenge communities was composed of new species led us to hypothesize that we could improve the colonization resistance of hCom1 by adding the invading species, thereby improving its ability to fill niches in the gut. Twenty-four bacterial species entered hCom1 from $\geq$2 of the 3 fecal samples used as a challenge (Table S4); we focused on these species, reasoning that they were more likely to fill conserved niches in the community. We were able to obtain 22/24 from culture collections, and we included all of them in the new community (hCom2). At the same time, we omitted seven species that either failed to colonize initially or were displaced in all three groups of mice (Table S4), reasoning that they were incompatible with the rest of hCom1 or incapable of colonizing the mouse gut under the dietary conditions in which the experiment was performed. Thus, the new community contains 97 strains from hCom1 plus 22 new strains, for a total of 119 (Figures 4A and S1; Table S2). These 22 strains are primarily Firmicutes or species of *Alistipes*. Many represent taxa that are phylogenetically under-represented in hCom1, suggesting that they might be able to occupy niches left open by the members of hCom1 (Figure S1).

We colonized four groups of germ-free SW mice with hCom2, collecting fecal pellets weekly (Figure 4A). As before, we measured community composition by analyzing metagenomic sequencing data with NinjaMap (Figure 5A; Table S4). The gut communities of hCom2-colonized mice rapidly reached a stable configuration (Pearson's r with respect to week 8 >0.97) (Figure S2). 100 of the 119 strains were above the limit of detection; hCom1-derived strains colonized at similar relative abundances in the context of the augmented community (with similarly low CVs across mice) (Figure 5B). The species that were new to hCom2 exhibited a wide range of relative abundances; *Bacteroides rodentium* became the most abundant species, whereas the least abundant of the new species, *Blautia* sp. KLE 1732, had a mean abundance $\sim 10^{-4}$ (Figure 5B).

The augmented community is more resilient to human fecal challenge

Our goal in constructing hCom2 was to improve its completeness as assessed by its ability to occupy niches in the gut. To test whether hCom2 is more complete than hCom1, we challenged hCom2-colonized mice at the beginning of week 5 with the same fecal samples used to challenge hCom1, enabling us to compare results between the challenge experiments. Importantly, the 22 strains used to augment hCom1 were obtained from culture collections rather than the fecal samples themselves, reducing the likelihood that hCom2 and the fecal samples have overlapping membership at the strain level (Garud et al., 2019). Indeed, by recruiting sequencing reads to the genomes of the new organisms in hCom2, we found that 17/22 were covered broadly ($\geq$95%) by perfectly matching reads, consistent with the view that they were derived from hCom2 and not the fecal challenge (Data S4).

An average of 96% of the genome copies (and 81% of the MIDAS bins) from week 8 derived from the strains in hCom2 (Figure 5C), demonstrating that the colonization resistance of hCom2 is markedly improved over hCom1 (Figure 5D). The remaining 4% of reads (and 19% of MIDAS bins) represent species that engrafted in the presence of hCom2 (Figures 5D and S2). Strikingly, nearly all of the species that invaded hCom2 also invaded hCom1 (Figure 5E; Table S4); we were either unable to obtain an isolate for inclusion in hCom2 or the species invaded



(legend on next page)

hCom1 from only 1 of the 3 fecal samples used as a challenge, falling below our threshold for inclusion. These species represented virtually all of the remaining genome copies. We conclude that more extensive augmentation, based on the results of the first challenge experiment, would likely have enhanced colonization resistance further.

Moreover, compared with hCom1, the composition of hCom2 post-challenge was more similar to its pre-challenge state (Pearson's r >0.95; Figure 5F). Taken together, these data show that hCom2 is more stable and complete than hCom1 and that the augmentation process is robust and fault tolerant in identifying species that can occupy unfilled niches.

In the previous experiment, we challenged hCom2-colonized mice with Hum1–3, the same fecal communities used in the initial augmentation experiment (Figure 4). We next sought to determine whether hCom2 is resilient to challenge by unrelated fecal communities. hCom2-colonized mice were challenged with Hum4-6, which are compositionally distinct from Hum1–3 (Figure 4A). hCom2 was somewhat less stable to challenge by unrelated fecal samples: an average of 81% of the genome copies from week 8 (and 58% of the MIDAS bins) derived from hCom2 (Figure 5D). Thus, hCom2 is broadly but not completely resilient to challenge by unrelated fecal samples.

The architecture of hCom2 resembles that of a complete, undefined human fecal consortium

Our original goal in building a complex defined community was to develop a model system for the gut microbiome. Having demonstrated that hCom2 is stable and resilient to invasion, we sought to assess whether it has the functional attributes of a model system.

We started by asking how its architecture—the relative abundances of its constituent taxa—compares with that of a human fecal community. We colonized germ-free mice with three human fecal samples (Hum1–3; hereafter, "humanized") and compared their community compositions to those of mice colonized with hCom2. The gut communities of hCom2-colonized and humanized mice were similar in three ways (Figures 5G, 5H, S3). First, relative abundances spanned at least five orders of magnitude, with some strains consistently colonizing at >10% and others at <0.001%. Second, the distribution of log relative abundances was centered at ~0.01%, indicating that the majority of strains in the community would be missed by enumeration tools that have a limit of detection of 0.1%. Third, relative abundances by taxon are similar down to the genus level (Figure S3). Thus, the architecture of hCom2 resembles that of a human fecal community in the mouse gut.

Reproducibility of colonization

We next addressed the question of reproducibility, which is a threshold requirement for an experimental model system. We started by analyzing data from the second fecal challenge experiment (with Hum1–3) to assess the technical reproducibility of community composition in mice colonized by hCom2. At week 4, strain abundances in 20 mice across 4 cages colonized by the same hCom2 inoculum were highly similar (pairwise Pearson's correlation coefficients 0.96 ± 0.01; Data S5).

Biological reproducibility was a greater concern. Given the complexity of hCom1 and hCom2, variability in the growth of individual strains could lead to substantial differences in the composition of inocula constructed on different days. To determine the extent to which this variability affects community architecture *in vivo*, we compared community composition in four groups of mice colonized by replicates of hCom2 constructed independently on different days (Figures 6A and 6B). The communities displayed a striking degree of similarity in relative abundance profiles after 4 weeks (Pearson's correlation coefficient >0.95 between all pairs of biological replicates). We conclude

Figure 5. An augmented community with improved resilience to fecal challenge

(A) Comparing the architecture and strain-level relative abundances of hCom1 and hCom2. Each column depicts the relative abundance of an individual strain from hCom2 across all samples at week 4. 100 of the 119 strains were detected; those that are new to hCom2 are colored red.

(B) Averaged relative abundances of the strains in hCom1 versus hCom2 at week 4. Strains that are new to hCom2 are indicated by a gray outline. Dots are colored by phylum according to the legend in (A).

(C) The architecture of hCom2 is largely unaffected by fecal challenge with Hum1–3. Each dot is an individual strain; the collection of dots in a column represents the community at a single time point averaged over the 5 co-housed mice that were challenged with fecal community Hum1. Strains are colored according to their rank-order relative abundance at week 4. Gray circles represent invading species, defined as any species not present in weeks 1–4 in the group of mice shown.

(D) Left: hCom2 is more resilient to fecal challenge than hCom1. Top row: blue squares in the waffle plots indicate MIDAS bins that derive from hCom2; gray squares represent MIDAS bins from the fecal communities. Bottom row: pie charts representing the percentage of MIDAS bins that derive from hCom2 versus the fecal communities. An average of 96% of the genome copies (and 81% of the MIDAS bins) come from hCom2 in the Hum1–3 challenges, demonstrating that the resilience of the community was improved markedly by augmentation with strains identified from the initial challenge (Figure 4). Right: hCom2 is broadly resilient to challenge by unrelated fecal samples (Hum4–6). In these challenges, an average of 81% of the genome copies (and 58% of the MIDAS bins) came from hCom2.

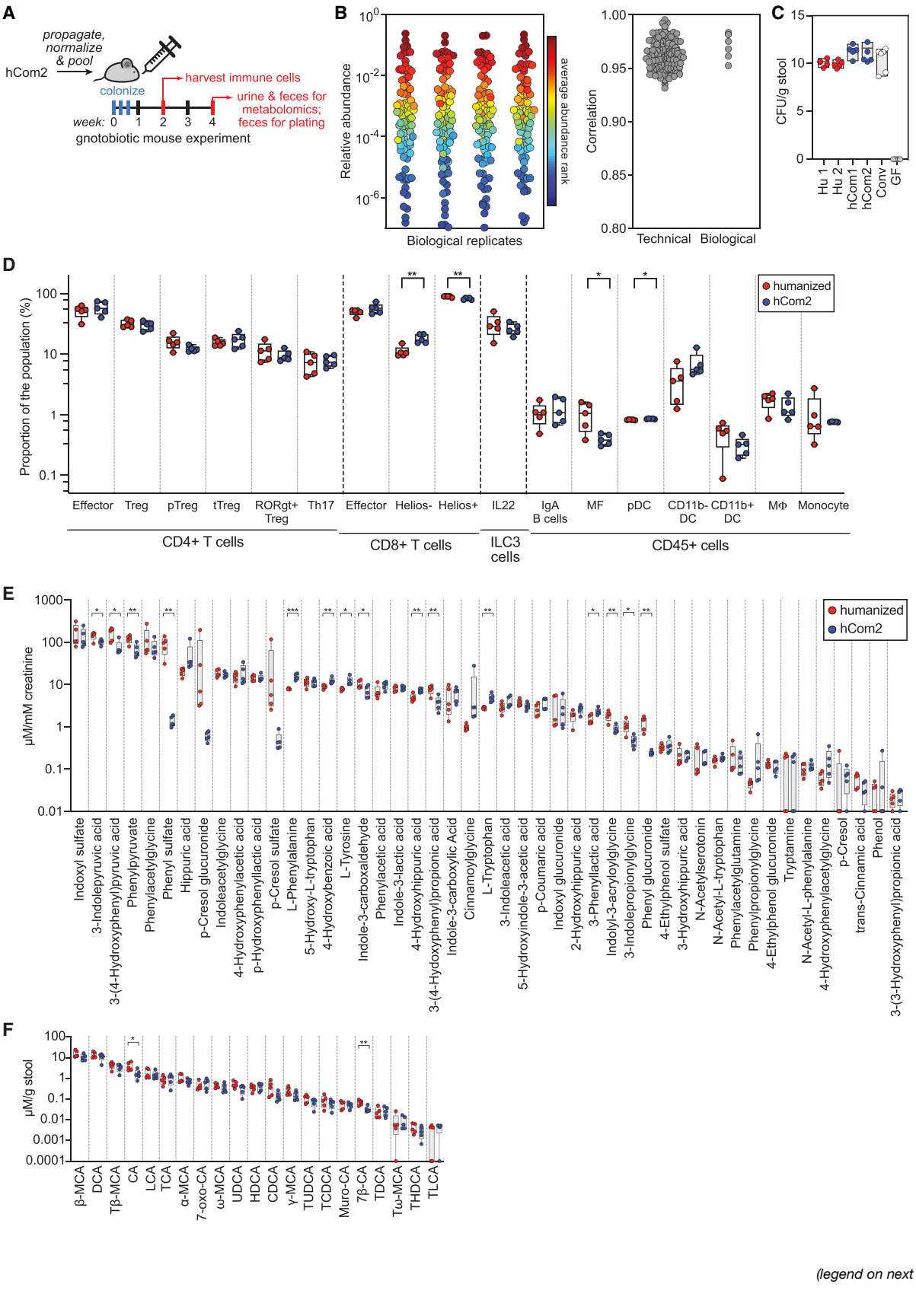
(E) Nearly all invading strains at week 8 were repeat invaders from the first fecal challenge (Table S4). The dots representing invading strains are shown in full color; dots representing hCom2-derived strains are partially transparent. Dots that represent repeat invaders from the first fecal challenge experiment have a thick black border.

(F) The relative abundances of the hCom2-derived species present post-challenge are highly correlated with their pre-challenge levels. Pearson's correlation coefficient with respect to the average relative abundance in weeks 3 and 4 are shown for the PBS control and 3 fecal community challenges, averaged across mice that received the same challenge. Correlation coefficients are shown for the 119 species in hCom2 (solid lines) and for all species including invaders (dashed lines).

(G) hCom2 resembles a fecal consortium more closely than hCom1. Averaged relative abundances of MIDAS bins are shown for hCom1- and hCom2-colonized mice versus mice colonized by a fecal community from one of three healthy human donors (Hum1–3). The phylum-level architecture of hCom2 is more closely correlated to that of humanized mice than hCom1 (Figure S3).

(H) Pairwise correlation coefficients of phylum-level relative abundance vectors were higher between hCom2-colonized and Hum1–3 humanized mice than between hCom1-colonized and Hum1–3 humanized mice.

See also Figures S1, S2, and S3, Tables S4 and S7, and Data S4.



(legend on next page)

that a relatively constant nutrient environment enables input communities with widely varying relative abundances to reach the same steady-state configuration, consistent with ecological observations in other microbial communities (Aranda-Díaz et al., 2022; Goldford et al., 2018; Hibberd et al., 2017; Venturelli et al., 2018). This high degree of biological reproducibility will be enabling for the use of complex defined communities as experimental models.

To further investigate the potential for hCom2 to function as a model microbiome, we assessed its composition in a second strain of mice. Since the experiments to develop hCom2 used outbred SW mice, we chose 129/SvEv, an inbred mouse strain. We colonized germ-free 129/SvEv mice with hCom2 and collected fecal pellets after 4 weeks of colonization. Community composition was highly correlated with that of SW mice (Pearson correlation coefficient > 0.95) (Data S5). These data indicate that hCom2, like the human gut microbiome (Rothschild et al., 2018), is robust to changes in host genotype.

hCom2-colonized mice are phenotypically similar to humanized mice

We performed three additional experiments to determine the degree to which hCom2-colonized mice resemble germ-free mice colonized by a human fecal community. Since our defined communities are composed of human fecal isolates, we colonized germ-free mice with hCom2 or an undefined human fecal community and assayed phenotypes after 4 weeks (Figure 6A). First, fecal pellets from each mouse were serially diluted and plated on Columbia blood agar to estimate the bacterial cell density in each community. Each group contained 10^{11} to 10^{12} colony forming units per gram of feces (Figure 6C), similar to previously reported estimates from humans and from conventional and humanized mice (Ley et al., 2006; Vandeputte et al., 2017). Thus, hCom2 colonizes the mouse gut to a similar density as a normal murine or human fecal community.

Next, we sought to determine whether mice colonized by hCom2 harbor a similar immune cell profile to that of humanized mice. We extracted and stained colonic immune cells and assayed them by flow cytometry. Most immune cell subtypes, including CD4+ T cells, IgA+ B cells, macrophages, CD11b+ dendritic cells, and monocytes, were similarly abundant in humanized and hCom2-colonized mice (Figure 6D; Data S5), indicating that—at least in broad terms—hCom2-colonized mice are immunologically comparable to humanized mice.

Finally, to determine whether hCom2-colonized and humanized mice harbor a similar profile of microbiome-derived metabolites, we analyzed fecal pellets and urine samples using targeted metabolomics. Aromatic amino acid metabolite levels in urine (Figure 6E) and primary and secondary bile acid levels in feces (Figure 6F) were comparable between hCom2-colonized and humanized mice. Taken together, these data suggest that hCom2 is a reasonable model of gut microbial metabolism.

hCom2 exhibits robust colonization resistance against pathogenic *Escherichia coli*

To demonstrate its utility as a model system, we used hCom2 to study an emergent property of gut communities: their ability to resist colonization by pathogens and pathobionts (Buffie et al., 2015). To test whether hCom2 exhibits colonization resistance, we studied invasion by *Escherichia coli* ATCC 43894, an enterohemorrhagic *E. coli* (EHEC). We chose this strain for three reasons. First, EHEC is responsible for life-threatening diarrheal infections and hemolytic uremic syndrome, and enteric colonization by other *E. coli* strains has been linked to malnutrition and inflammatory bowel disease (Palmela et al., 2018; Pham et al., 2019). Second, colonization resistance to *E. coli* and other Enterobacteriaceae has been studied in detail (Litvak et al., 2019; Stromberg et al., 2018; Velazquez et al., 2019), but the commensal strains responsible and mechanisms by which they act are incompletely understood. Finally, hCom2 harbors no Enterobacteriaceae and only three species of Proteobacteria (*Desulfovibrio piger*, *Bilophila wadsworthia*, and *Burkholderiales bacterium* 1-1-47); hence, resistance to *E. coli* colonization would require a mechanism other than exclusion by a close relative occupying the same niche.

To test whether hCom2 is capable of resisting EHEC engraftment, we colonized germ-free SW mice with hCom2 or one of two other communities: a 12-member community (12Com) similar to one used in previous studies (McNulty et al., 2013) or an undefined fecal community from a healthy human donor (Figure 7A). hCom2 and 12Com do not contain any Enterobacteriaceae. To test whether non-pathogenic Enterobacteriaceae

Figure 6. hCom2-colonized mice are phenotypically similar to humanized mice

(A) Schematic of the experiment. Germ-free SW mice were colonized with freshly prepared hCom2 or a fecal sample from a healthy human donor. One cohort of mice was sacrificed at 2 weeks for immune cell profiling; another was sacrificed at 4 weeks for targeted metabolite analysis.

(B) The architecture of hCom2 in mice is highly reproducible. Left: community composition is highly similar across four biological replicates. Each dot is an individual strain; the collection of dots in a column represents the community at 4 weeks averaged over 5 mice co-housed in a cage. Strains are colored according to their average rank-order relative abundance across all samples. Right: Pearson's pairwise correlation coefficients for technical and biological replicates.

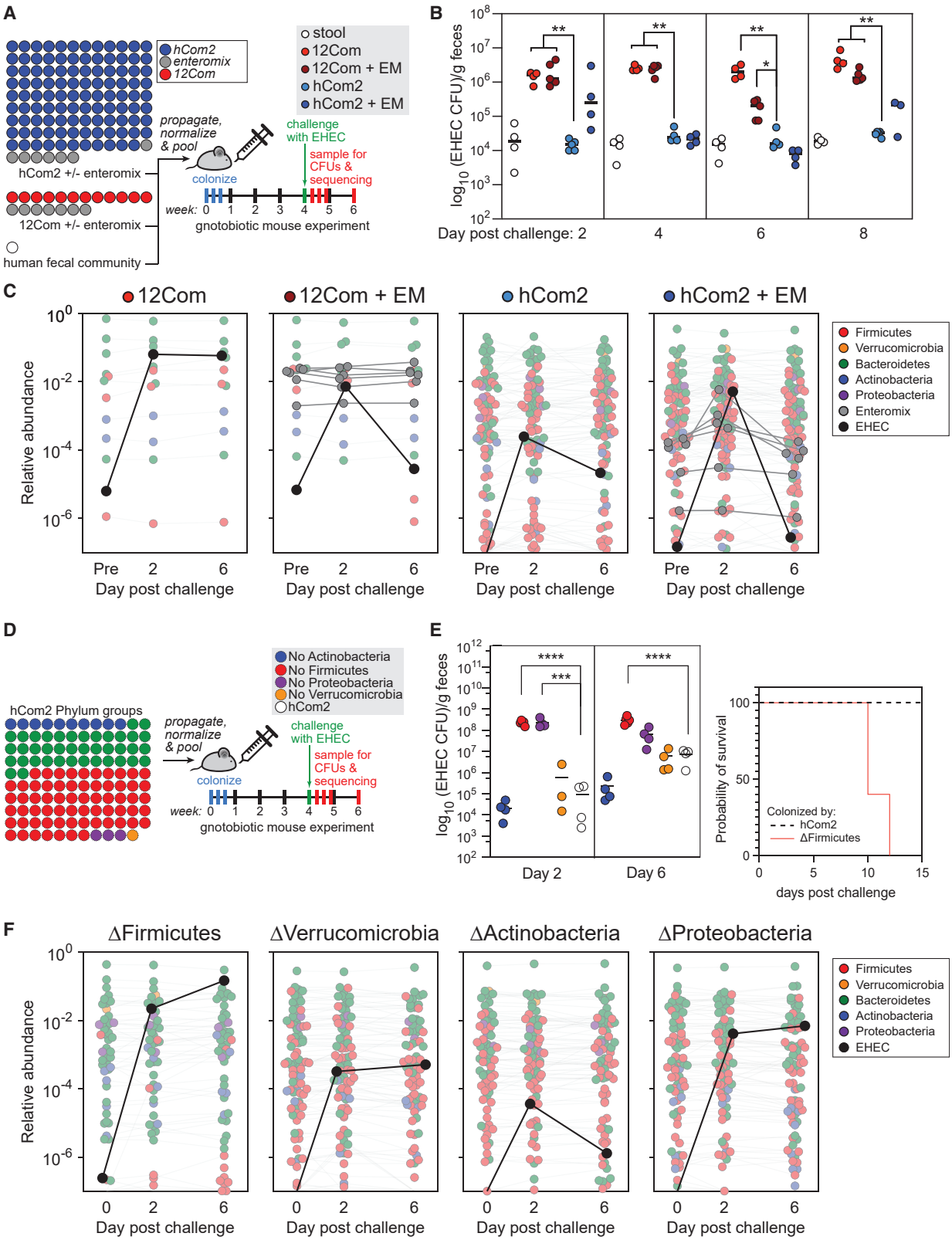
(C) hCom2-colonized, hCom1-colonized, and humanized mice have similar bacterial cell densities *in vivo*. Fecal samples from hCom2-colonized, hCom1-colonized, humanized, specific pathogen-free (SPF), or germ-free (GF) mice were homogenized and plated anaerobically on Columbia blood agar to enumerate colony forming units.

(D) Immune cell types and numbers were broadly similar between hCom2-colonized and humanized mice. Colonic immune cells were extracted from hCom2-colonized, humanized, or germ-free mice (all C57BL/6), stained for cell surface markers, and assessed by flow cytometry. Statistical significance was assessed using a Student's two tailed t test (**$p < 0.05$).

(E) hCom2-colonized mice and humanized mice have a similar profile of microbiome-derived metabolites. Urine samples from hCom2-colonized and humanized mice were analyzed by targeted metabolomics to measure a panel of aromatic amino acid metabolites by LC-MS. Statistical significance was assessed using a Student's two tailed t test (*$p < 0.05$ and **$p < 0.001$).

(F) Bile acids were extracted from fecal pellets collected from hCom2-colonized and humanized mice and were quantified by LC-MS. Statistical significance was assessed using a Student's two tailed t test (*$p < 0.05$ and **$p < 0.001$).

See also Data S5.



(legend on next page)

enhance colonization resistance to EHEC, we colonized two additional groups of mice with variants of hCom2 and 12Com to which a mixture of seven non-pathogenic Enterobacteriaceae strains were added (*Escherichia coli* MITI 27, *Escherichia coli* MITI 117, *Escherichia coli* MITI 135, *Escherichia coli* MITI 139, *Escherichia coli* MITI 255, *Escherichia coli* MITI 284, and *Enterobacter cloacae* MITI 173; termed "Enteromix"). After 4 weeks, we challenged with EHEC and assessed invasion by selective plating under aerobic growth conditions (Figure 7A).

Consistent with previous reports (Mohawk and O'Brien, 2011; Stromberg et al., 2018), the undefined human fecal community conferred robust resistance against EHEC colonization (Figures 7B and 7C). In contrast, 12Com allowed much higher levels of EHEC growth; the addition of Enteromix to 12Com improved the phenotype but did not restore full EHEC resistance (Figure 7B). Despite lacking Enterobacteriaceae, hCom2 exhibited a similar level of EHEC resistance to that of an undefined fecal community (Figure 7B). Thus, hCom2 is sufficiently complete to exhibit comparable levels of colonization resistance with a native fecal community.

As a starting point for identifying which species in hCom2 are responsible for EHEC colonization resistance, we constructed four communities in which we dropped out, in turn, all of the species in the phyla Firmicutes, Verrucomicrobia, Actinobacteria, and Proteobacteria. We colonized mice with these phylum dropout communities and then challenged them with EHEC (Figure 7D). The ΔActinobacteria (missing 10 strains) and ΔVerrucomicrobia communities (missing 1 strain, *Akkermansia muciniphila*) resisted EHEC comparably with hCom2 (Figures 7E and 7F). However, the ΔProteobacteria and ΔFirmicutes communities were more susceptible. Thus, despite the lack of Enterobacteriaceae in hCom2, the absence of the three more distantly related species of Proteobacteria was sufficient to confer sensitivity to EHEC invasion.

The ΔFirmicutes community was highly sensitive to EHEC invasion (Figure 7E); the defect resulted in a large survival difference between hCom2-colonized and Δ*Firmicutes*-colonized mice (Figure 7E, right). These results indicate either that either Firmicutes play a role in EHEC resistance or that a change in community architecture induced by their removal renders the community sensitive to invasion. Further studies with more precise strain dropout experiments could uncover strains that confer resistance and may enable more targeted microbial therapy against EHEC colonization and infection.

DISCUSSION

By developing a community that is both defined and reasonably complex, we have generated a model system that captures much of the biology of a native microbiome. Future refinements are needed, including additional bacterial strains to occupy unfilled niches as well as archaea, fungi, and viruses, all of which are important components of the native ecosystem.

The computational pipeline we developed for read mapping makes it possible to analyze complex defined communities with high precision and sensitivity. Community structure can be quantified across six orders of magnitude in relative abundance, enabling the interrogation of low-abundance community members that play important roles in community function and dynamics (Buffie et al., 2015; Funabashi et al., 2020). The degree of technical and biological reproducibility (Figure 6B) is remarkable in a system this complex, which bodes well for future experimental efforts.

The process by which we augmented a defined community revealed two unexpected findings. First, a community composed of strains from >100 distinct donors can be stable *in vivo*. It remains to be seen whether there are appreciable differences in stability—or in fine-scale genomic and phenotypic adaptation—between communities composed of isolates from a single donor (in which strains have co-existed for years) versus multiple donors (in which strains have no prior history together). If a collection of strains with no common history can form a stable consortium, it will be interesting to determine the role of priority

Figure 7. hCom2 exhibits colonization resistance against enterohemorrhagic *E. coli*

(A) Schematic of the experiment. We colonized germ-free SW mice with freshly prepared hCom2 or one of two other communities: a 12-member synthetic community (12Com) or a fecal community from a healthy human donor. hCom2 and 12Com do not contain any Enterobacteriaceae; to test whether non-pathogenic Enterobacteriaceae enhance colonization resistance to EHEC, we colonized two additional groups of mice with variants of hCom2 and 12Com to which a mixture of seven non-pathogenic Enterobacteriaceae strains were added (six *E. coli* and *Enterobacter cloacae*, Enteromix [EM]). After 4 weeks, we challenged with 10^9 colony forming units of EHEC and assessed the degree to which it colonized in two ways: by EHEC-selective plating under aerobic growth conditions and by metagenomic sequencing with NinjaMap analysis.

(B) hCom2 exhibits a similar degree of EHEC resistance to that of a fecal community in mice. Colony forming units of EHEC in mice colonized by the four communities are shown. As expected, the fecal community conferred robust colonization resistance while 12Com did not. The addition of EM moderately improved the EHEC resistance of 12Com. Despite lacking Enterobacteriaceae, hCom2 exhibited a similar level of EHEC resistance to that of an undefined fecal community.

(C) The architecture of hCom2 is stable following EHEC challenge. Each dot is an individual strain; the collection of dots in a column represents the community at a single time point averaged over four co-housed mice. Strains are colored according to their phylum; EHEC is shown in black and members of the Enteromix community are shown in gray.

(D) Schematic of the phylum dropout experiment. We colonized germ-free SW mice with four variants of hCom2, each one missing all species from the phyla Actinobacteria, Firmicutes, Proteobacteria, or Verrucomicrobia. After 4 weeks, we challenged with 10^9 colony forming units of EHEC and assessed the degree to which it colonized by EHEC-selective plating under aerobic growth conditions and by metagenomic sequencing with NinjaMap analysis.

(E) The ΔActinobacteria and ΔVerrucomicrobia communities retain the ability to resist EHEC invasion, whereas the ΔFirmicutes and ΔProteobacteria communities are sensitive to EHEC invasion. Right: a large survival difference in ΔFirmicutes-colonized mice compared with hCom2-colonized mice.

(F) The architecture of the phylum dropout communities remains stable following EHEC challenge. Each dot is an individual strain; the collection of dots in a column represents the community at a single time point averaged over four co-housed mice. Strains are colored according to their phylum; EHEC is shown in black.

effects (i.e., order of arrival) and spatial and metabolic niche occupancy.

Second, the process we introduce here for filling open niches is surprisingly robust and fault tolerant. Most notably, nearly all of the fecal community-derived strains that invaded hCom1—*Alistipes*, *Blautia*, *Bilophila*, *Oscilibacter*, and Proteobacteria—were under-represented phylogenetically within hCom1 (Figure S1). Moreover, most of the strains that invaded hCom2 had previously invaded hCom1, indicating that niche filling is deterministic. Importantly, the augmentation process caused relatively little perturbation to the structure of the existing community (notable exceptions are shown in Table S4), suggesting that it will result in a progressive improvement of the community. Although the augmentation process can only fill niches that are conserved from mice to humans, the observation that most of our human strains engrafted suggests that many niches are conserved.

If we had broadened our strain inclusion criteria, there is a reasonable likelihood we could have improved colonization resistance further after just one round of augmentation. To further enhance niche filling and stability, it would help to subject hCom2 to further rounds of augmentation using fecal samples from additional donors, ideally in the presence of a varying diet. It might also be possible to improve niche occupancy, for example, in the setting of intestinal inflammation by performing the augmentation process in a murine model of inflammatory bowel disease.

There is a pressing need for a common model system for the gut microbiome that is completely defined and complex enough to capture much of the biology of a full-scale community. We showed that hCom2 is a reasonable starting point for such a system: in spite of its complexity, it colonizes mice in a highly reproducible manner. Moreover, hCom2 faithfully models the carrying capacity, immune cell profile, and metabolic phenotypes of humanized mice. There remain some modest differences in metabolic and immune profiles, and the community is still missing certain taxa that will likely be important to add. Nonetheless, taken together, our findings suggest that hCom2 is a reasonable starting point for a model of the gut microbiome.

One of the most interesting possibilities for such a system would be to enable reductionist experiments downstream of a community transplantation experiment (e.g., to identify strains responsible for a microbiome-linked phenotype). Although we did not identify the strains responsible for colonization resistance to EHEC, we did find that removing species of Proteobacteria or Firmicutes rendered the community EHEC-sensitive. Follow-up experiments in which one or several strains at a time are eliminated from the community could narrow further from the phylum level to individual strains. Efforts to identify the strains responsible for other microbiome-linked phenotypes, including response to cancer immunotherapy, caloric harvest, and neural development, would be of great interest.

Limitations of the study

Our study has three important limitations. First, although hCom2 is stable to challenge with the fecal communities used to augment it, it is less stable to challenge with unrelated fecal communities. These data suggest that subsequent rounds of backfill—using a variety of unrelated fecal samples in series or in parallel—is a promising path toward an even stabler variant of hCom2.

Second, it is unclear how many more bacterial strains (or other components) may be necessary to model the full functional capacity of a native human microbiome. Prior estimates of the number of species in a typical human microbiome range from ~150–300 (Faith et al., 2013; Kraal et al., 2014; Qin et al., 2010). Nonetheless, the observation that a defined community of just 119 strains exhibits remarkable stability bodes well for future efforts. We estimate that hCom2 is within 2-fold of native-scale complexity (STAR Methods), so a full-scale system is experimentally feasible. As a starting point for efforts to build such a system, hCom2 will provide a standard for assessing the genomic and functional completeness of model communities, with the ultimate goal of modeling native-scale human microbiomes.

Third, strain-level variation among communities underlies some of the phenotypic differences conferred on the host by the microbiome (Campbell et al., 2020; Jin et al., 2022; Marcobal et al., 2011; McNulty et al., 2011). hCom2 represents just one consortium of strains; therefore, neither hCom2 nor any other single community can model the impact of strain-level variation on host phenotype. However, we think that a defined community is a promising starting point for probing strain-level differences: a collection of communities that are identical but harbor different strains of a species of interest would be an ideal way to probe the impact of strain variation—or even individual genes—on phenotype.

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 SUBJECT DETAILS
  - Bacterial strains and culture conditions
  - Preparation of synthetic community for storage and for experiments
  - Preparation of synthetic community for *in vitro* experiments
  - Preparation of synthetic community for *in vivo* experiments
  - Collection and preservation of human fecal samples
  - Preparation of human fecal samples
  - Preparation of 12Com
  - Preparation of Enteromix
- METHOD DETAILS
  - Metagenomic sequencing
  - Constructing high quality genome assemblies
  - Generating and normalizing the NinjaMap database
  - Generating simulated sequencing reads
  - Augmenting the NinjaMap database

- Metagenomic read mapping
- Sensitivity of NinjaMap
- Amino acid dropout experiment and data analysis
- Constructing *C. sporogenes* mutants
- ATP assay
- Reproducibility and colonization experiments
- Augmentation experiment
- Bacterial load estimates
- Immune profiling
- Metabolomics
- *E. coli* colonization resistance
- Estimation that hCom2 is within two-fold of native-scale complexity
- A compilation of estimates from the literature

● QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2022.08.003.

ACKNOWLEDGMENTS

We are deeply indebted to members of the Fischbach and Huang labs for helpful discussions and to Rod Mackie, University of Illinois at Urbana-Champaign (UIUC), for bacterial strains used in this study. A.A.-D. is a Howard Hughes Medical Institute International Student Research fellow, a Stanford Bio-X Bowes fellow, and a Siebel Scholar. This work was supported by a Dean's Postdoctoral Fellowship (to P.-Y.H.), NIH F32GM143859 (to P.-Y.H.), Human Frontier Science Program award LT000493/2018-L (to K.N.), a Fellowship from the Astellas Foundation for Research on Metabolic Disorders (to K.N.), the Stanford Microbiome Therapies Initiative (to M.A.F. and K.C.H.), NIH grants DP1 DK113598 (to M.A.F.), P01 HL147823 (to M.A.F.), R01 DK101674 (to M.A.F. and J.L.S.), RM1 GM135102 (to K.C.H.), and R01 AI147023 (to K.C.H.), NSF grant EF-2125383 (to K.C.H. and M.A.F.), the Helmsley Charitable Trust (to M.A.F.), the Bill and Melinda Gates Foundation (to M.A.F.), an HHMI-Simons Faculty Scholars Award (to M.A.F.), the Leducq Foundation (to M.A.F.), the Stanford-Coulter Translational Research Grants Program (to M.A.F.), MAC3 Impact Philanthropies (to M.A.F.), and the Allen Discovery Center at Stanford on Systems Modeling of Infection (to K.C.H.). K.C.H., J.L.S., and M.A.F. are Chan Zuckerberg Biohub investigators.

AUTHOR CONTRIBUTIONS

Conceptualization, A.G.C. and M.A.F.; methodology and investigation, A.G.C., P.-Y.H., A.A.-D. S.J., F.B.Y., X.M., M.W., M.I., K.N., A.Z., P.M., A.P., K.A., A.W., J.Y., A.R.B., S.H., A.D., A.L.S., A.D., N.N., J.L.S., K.C.H., and M.A.F.; visualization, A.G.C., P.-Y.H., A.A.-D., S.J., K.C.H., and M.A.F.; supervision, N.N., J.L.S., K.C.H., and M.A.F.; writing, A.G.C., P.-Y.H., A.A.-D., S.J., K.C.H., and M.A.F. All authors reviewed the manuscript before submission.

DECLARATION OF INTERESTS

Stanford University and the Chan Zuckerberg Biohub have patents pending for microbiome technologies on which the authors are co-inventors. M.A.F. is a co-founder and director of Federation Bio and Kelonia, a co-founder of Revolution Medicines, and a member of the scientific advisory boards of NGM Bio and Zymergen. A.G.C. and K.N. have been paid consultants to Federation Bio. A.R.B. has been an employee of Federation Bio.

Received: June 9, 2021
Revised: March 2, 2022
Accepted: August 3, 2022
Published: September 6, 2022

REFERENCES

Angly, F.E., Willner, D., Rohwer, F., Hugenholtz, P., and Tyson, G.W. (2012). Grinder: a versatile amplicon and shotgun sequence simulator. Nucleic Acids Res. *40*, e94.

Aranda-Díaz, A., Ng, K.M., Thomsen, T., Real-Ramírez, I., Dahan, D., Dittmar, S., Gonzalez, C.G., Chavez, T., Vasquez, K.S., Nguyen, T.H., et al. (2022). Establishment and characterization of stable, diverse, fecal-derived in vitro microbial communities that model the intestinal microbiota. Cell Host & Microbe *30*. https://doi.org/10.1016/j.chom.2021.12.008.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. *19*, 455–477.

Blasche, S., Kim, Y., Oliveira, A.P., and Patil, K.R. (2017). Model microbial communities for ecosystems biology. Curr. Opin. Syst. Biol. *6*, 51–57.

Buffie, C.G., Bucci, V., Stein, R.R., McKenney, P.T., Ling, L., Gobourne, A., No, D., Liu, H., Kinnebrew, M., Viale, A., et al. (2015). Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. Nature *517*, 205–208.

Buffie, C.G., and Pamer, E.G. (2013). Microbiota-mediated colonization resistance against intestinal pathogens. Nat. Rev. Immunol. *13*, 790–801.

Buffington, S.A., Dooling, S.W., Sgritta, M., Noecker, C., Murillo, O.D., Felice, D.F., Turnbaugh, P.J., and Costa-Mattioli, M. (2021). Dissecting the contribution of host genetics and the microbiome in complex behaviors. Cell *184*, 1740–1756.e16.

Campbell, C., McKenney, P.T., Konstantinovsky, D., Isaeva, O.I., Schizas, M., Verter, J., Mai, C., Jin, W.-B., Guo, C.-J., Violante, S., et al. (2020). Bacterial metabolism of bile acids promotes generation of peripheral regulatory T cells. Nature *581*, 475–479.

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.

Cunin, R., Glansdorff, N., Piérard, A., and Stalon, V. (1986). Biosynthesis and metabolism of arginine in bacteria. Microbiol. Rev. *50*, 314–352.

Deschasaux, M., Bouter, K.E., Prodan, A., Levin, E., Groen, A.K., Herrema, H., Tremaroli, V., Bakker, G.J., Attaye, I., Pinto-Sietsma, S.-J., et al. (2018). Depicting the composition of gut microbiota in a population with varied ethnic origins but shared geography. Nat. Med. *24*, 1526–1531.

Dethlefsen, L., and Relman, D.A. (2011). Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. Proc. Natl. Acad. Sci. USA *108*, 4554–4561.

Dodd, D., Spitzer, M.H., Van Treuren, W., Merrill, B.D., Hryckowian, A.J., Higginbottom, S.K., Le, A., Cowan, T.M., Nolan, G.P., Fischbach, M.A., and Sonnenburg, J.L. (2017). A gut bacterial pathway metabolizes aromatic amino acids into nine circulating metabolites. Nature *551*, 648–652.

Faith, J.J., Ahern, P.P., Ridaura, V.K., Cheng, J., and Gordon, J.I. (2014). Identifying gut microbe-host phenotype relationships using combinatorial communities in gnotobiotic mice. Sci. Transl. Med. *6*, 220ra11.

Faith, J.J., Guruge, J.L., Charbonneau, M., Subramanian, S., Seedorf, H., Goodman, A.L., Clemente, J.C., Knight, R., Heath, A.C., Leibel, R.L., et al. (2013). The long-term stability of the human gut microbiota. Science *341*, 1237439.

Faith, J.J., McNulty, N.P., Rey, F.E., and Gordon, J.I. (2011). Predicting a human gut microbiota's response to diet in gnotobiotic mice. Science *333*, 101–104.

Franzosa, E.A., Huang, K., Meadow, J.F., Gevers, D., Lemon, K.P., Bohannan, B.J.M., and Huttenhower, C. (2015). Identifying personal microbiomes using metagenomic codes. Proc. Natl. Acad. Sci. USA *112*, E2930–E2938.

Funabashi, M., Grove, T.L., Wang, M., Varma, Y., McFadden, M.E., Brown, L.C., Guo, C., Higginbottom, S., Almo, S.C., and Fischbach, M.A. (2020). A metabolic pathway for bile acid dehydroxylation by the gut microbiome. Nature *582*, 566–570.

Garud, N.R., Good, B.H., Hallatschek, O., and Pollard, K.S. (2019). Evolutionary dynamics of bacteria in the gut microbiome within and across hosts. PLoS Biol *17*. e3000102. https://doi.org/10.1371/journal.pbio.3000102.

Goldford, J.E., Lu, N., Bajić, D., Estrela, S., Tikhonov, M., Sanchez-Gorostiaga, A., Segrè, D., Mehta, P., and Sanchez, A. (2018). Emergent simplicity in microbial community assembly. Science *361*, 469–474.

Goodman, A.L., Kallstrom, G., Faith, J.J., Reyes, A., Moore, A., Dantas, G., and Gordon, J.I. (2011). Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. Proc. Natl. Acad. Sci. USA *108*, 6252–6257.

Goodman, A.L., McNulty, N.P., Zhao, Y., Leip, D., Mitra, R.D., Lozupone, C.A., Knight, R., and Gordon, J.I. (2009). Identifying genetic determinants needed to establish a human gut symbiont in its habitat. Cell Host Microbe *6*, 279–289.

Gopalakrishnan, V., Spencer, C.N., Nezi, L., Reuben, A., Andrews, M.C., Karpinets, T.V., Prieto, P.A., Vicente, D., Hoffman, K., Wei, S.C., et al. (2018). Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. Science *359*, 97–103.

Guarner, F., and Malagelada, J.-R. (2003). Gut flora in health and disease. Lancet *361*, 512–519.

Guo, C.-J., Allen, B.M., Hiam, K.J., Dodd, D., Van Treuren, W., Higginbottom, S., Nagashima, K., Fischer, C.R., Sonnenburg, J.L., Spitzer, M.H., and Fischbach, M.A. (2019). Depletion of microbiome-derived molecules in the host using *Clostridium* genetics. Science *366*, eaav1282.

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. Bioinformatics *29*, 1072–1075.

He, Y., Wu, W., Zheng, H.-M., Li, P., McDonald, D., Sheng, H.-F., Chen, M.-X., Chen, Z.-H., Ji, G.-Y., Zheng, Z.-D.-X., et al. (2018). Regional variation limits applications of healthy gut microbiome reference ranges and disease models. Nat. Med. *24*, 1532–1535.

Hibberd, M.C., Wu, M., Rodionov, D.A., Li, X., Cheng, J., Griffin, N.W., Barratt, M.J., Giannone, R.J., Hettich, R.L., Osterman, A.L., and Gordon, J.I. (2017). The effects of micronutrient deficiencies on bacterial species from the human gut microbiota. Sci. Transl. Med. *9*.

Holdeman, L.V. (1975). Discussion of Current Bacteriological Investigations of the Relationships between Intestinal Flora, Diet, and Colon Cancer. Cancer Research *35*, 3418–3420.

Jin, W.-B., Li, T.-T., Huo, D., Qu, S., Li, X.V., Arifuzzaman, M., Lima, S.F., Shi, H.-Q., Wang, A., Putzel, G.G., et al. (2022). Genetic manipulation of gut microbes enables single-gene interrogation in a complex microbiome. Cell *185*, 547–562. e22.

Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., and Wang, Z. (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ *7*, e7359.

Kraal, L., Abubucker, S., Kota, K., Fischbach, M.A., and Mitreva, M. (2014). The prevalence of species and strains in the human microbiome: a resource for experimental efforts. PLoS One *9*, e97279.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lawley, T.D., and Walker, A.W. (2013). Intestinal colonization resistance. Immunology *138*, 1–11.

Ley, R.E., Peterson, D.A., and Gordon, J.I. (2006). Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell *124*, 837–848.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079.

Litvak, Y., Mon, K.K.Z., Nguyen, H., Chanthavixay, G., Liou, M., Velazquez, E.M., Kutter, L., Alcantara, M.A., Byndloss, M.X., Tiffany, C.R., et al. (2019). Commensal Enterobacteriaceae protect against *Salmonella* colonization through oxygen competition. Cell Host Microbe *25*, 128–139.e5.

Lu, J., Breitwieser, F.P., Thielen, P., and Salzberg, S.L. (2017). Bracken: estimating species abundance in metagenomics data. PeerJ Comput. Sci. *3*, e104.

Marcobal, A., Barboza, M., Sonnenburg, E.D., Pudlo, N., Martens, E.C., Desai, P., Lebrilla, C.B., Weimer, B.C., Mills, D.A., German, J.B., and Sonnenburg, J.L. (2011). *Bacteroides in the infant gut consume milk oligosaccharides via mucus-utilization pathways*. Cell Host Microbe *10*, 507–514.

Martens, E.C., Kelly, A.G., Tauzin, A.S., and Brumer, H. (2014). The devil lies in the details: how variations in polysaccharide fine-structure impact the physiology and evolution of gut microbes. J. Mol. Biol. *426*, 3851–3865.

Matson, V., Fessler, J., Bao, R., Chongsuwat, T., Zha, Y., Alegre, M.-L., Luke, J.J., and Gajewski, T.F. (2018). The commensal microbiome is associated with anti-PD-1 efficacy in metastatic melanoma patients. Science *359*, 104–108.

McNulty, N.P., Wu, M., Erickson, A.R., Pan, C., Erickson, B.K., Martens, E.C., Pudlo, N.A., Muegge, B.D., Henrissat, B., Hettich, R.L., and Gordon, J.I. (2013). *Effects of diet on resource utilization by a model human gut microbiota containing Bacteroides cellulosilyticus WH2, a symbiont with an extensive glycobiome*. PLoS Biol. *11*, e1001637.

McNulty, N.P., Yatsunenko, T., Hsiao, A., Faith, J.J., Muegge, B.D., Goodman, A.L., Henrissat, B., Oozeer, R., Cools-Portier, S., Gobert, G., et al. (2011). The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins. Sci. Transl. Med. *3*, 106ra106.

Mohawk, K.L., and O'Brien, A.D. (2011). Mouse models of *Escherichia coli* O157:H7 infection and shiga toxin injection. J. Biomed. Biotechnol. *2011*, 258185.

Morris, B.E.L., Henneberger, R., Huber, H., and Moissl-Eichinger, C. (2013). Microbial syntrophy: interaction for the common good. FEMS Microbiol. Rev. *37*, 384–406.

Nayfach, S., Rodriguez-Mueller, B., Garud, N., and Pollard, K.S. (2016). An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. Genome Res. *26*, 1612–1625.

Nayfach, S., Shi, Z.J., Seshadri, R., Pollard, K.S., and Kyrpides, N.C. (2019). New insights from uncultivated genomes of the global human gut microbiome. Nature *568*, 505–510.

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.

Nisman, B. (1954). The Stickland reaction. Bacteriol. Rev. *18*, 16–42.

O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. *44*, D733–D745.

Pacheco, A.R., and Segrè, D. (2019). A multidimensional perspective on microbial interactions. FEMS Microbiol. Lett. *366*, fnz125.

Palmela, C., Chevarin, C., Xu, Z., Torres, J., Sevrin, G., Hirten, R., Barnich, N., Ng, S.C., and Colombel, J.-F. (2018). Adherent-invasive *Escherichia coli* in inflammatory bowel disease. Gut *67*, 574–587.

Parks, D.H., Chuvochina, M., Chaumeil, P.-A., Rinke, C., Mussig, A.J., and Hugenholtz, P. (2020). A complete domain-to-species taxonomy for Bacteria and Archaea. Nat. Biotechnol. *38*, 1079–1086.

Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., and Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat. Biotechnol. *36*, 996–1004.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. *25*, 1043–1055.

Patnode, M.L., Beller, Z.W., Han, N.D., Cheng, J., Peters, S.L., Terrapon, N., Henrissat, B., Le Gall, S., Saulnier, L., Hayashi, D.K., et al. (2019). Interspecies competition impacts targeted manipulation of human gut bacteria by fiber-derived glycans. Cell *179*, 59–73.e13.

Pham, T.-P.-T., Tidjani Alou, M., Bachar, D., Levasseur, A., Brah, S., Alhousseini, D., Sokhna, C., Diallo, A., Wieringa, F., Million, M., et al. (2019). Gut microbiota alteration is characterized by a proteobacteria and fusobacteria

bloom in kwashiorkor and a bacteroidetes paucity in marasmus. Sci. Rep. *9*, 9084.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. Nature *464*, 59–65.

Qin, M., Wu, S., Li, A., Zhao, F., Feng, H., Ding, L., Chang, Y., and Ruan, J. (2018). LRScaf: improving draft genomes using long noisy reads. Preprint at bioRxiv. https://doi.org/10.1101/374868.

Ridaura, V.K., Faith, J.J., Rey, F.E., Cheng, J., Duncan, A.E., Kau, A.L., Griffin, N.W., Lombard, V., Henrissat, B., Bain, J.R., et al. (2013). Gut microbiota from twins discordant for obesity modulate metabolism in mice. Science *341*, 1241214.

Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., Costea, P.I., Godneva, A., Kalka, I.N., Bar, N., et al. (2018). Environment dominates over host genetics in shaping human gut microbiota. Nature *555*, 210–215.

Routy, B., Le Chatelier, E., Derosa, L., Duong, C.P.M., Alou, M.T., Daillère, R., Fluckiger, A., Messaoudene, M., Rauber, C., Roberti, M.P., et al. (2018). Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. Science *359*, 91–97.

Sharon, G., Cruz, N.J., Kang, D.-W., Gandal, M.J., Wang, B., Kim, Y.-M., Zink, E.M., Casey, C.P., Taylor, B.C., Lane, C.J., et al. (2019). Human gut microbiota from autism spectrum disorder promote behavioral symptoms in mice. Cell *177*, 1600–1618.e17.

Shen, W., Le, S., Li, Y., and Hu, F. (2016). SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. PLoS One *11*, e0163962.

Smith, E.A., and Macfarlane, G.T. (1997). Dissimilatory amino acid metabolism in human colonic bacteria. Anaerobe *3*, 327–337.

Sonnenburg, E.D., and Sonnenburg, J.L. (2019). The ancestral and industrialized gut microbiota and implications for human health. Nat. Rev. Microbiol. *17*, 383–390.

Soto-Martin, E.C., Warnke, I., Farquharson, F.M., Christodoulou, M., Horgan, G., Derrien, M., Faurie, J.-M., Flint, H.J., Duncan, S.H., and Louis, P. (2020). Vitamin biosynthesis by human gut butyrate-producing bacteria and cross-feeding in synthetic microbial communities. mBio *11*. e00886–20.

Stromberg, Z.R., Van Goor, A., Redweik, G.A.J., Wymore Brand, M.J., Wannemuehler, M.J., and Mellata, M. (2018). Pathogenic and non-pathogenic *Escherichia coli* colonization and host inflammatory response in a defined microbiota mouse model. Dis. Model. Mech. *11*. dmm035063.

Titus Brown, C., and Irber, L. (2016). sourmash: a library for MinHash sketching of DNA. JOSS *1*, 27.

Truong, D.T., Franzosa, E.A., Tickle, T.L., Scholz, M., Weingart, G., Pasolli, E., Tett, A., Huttenhower, C., and Segata, N. (2015). MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nat. Methods *12*, 902–903.

van der Lelie, D., Oka, A., Taghavi, S., Umeno, J., Fan, T.-J., Merrell, K.E., Watson, S.D., Ouellette, L., Liu, B., Awoniyi, M., et al. (2021). Rationally designed bacterial consortia to treat chronic immune-mediated colitis and restore intestinal homeostasis. Nat. Commun. *12*, 3105.

Vandeputte, D., Kathagen, G., D'hoe, K., Vieira-Silva, S., Valles-Colomer, M., Sabino, J., Wang, J., Tito, R.Y., De Commer, L., Darzi, Y., et al. (2017). Quantitative microbiome profiling links gut community variation to microbial load. Nature *551*, 507–511.

Velazquez, E.M., Nguyen, H., Heasley, K.T., Saechao, C.H., Gil, L.M., Rogers, A.W.L., Miller, B.M., Rolston, M.R., Lopez, C.A., Litvak, Y., et al. (2019). Endogenous Enterobacteriaceae underlie variation in susceptibility to *Salmonella* infection. Nat. Microbiol. *4*, 1057–1064.

Venturelli, O.S., Carr, A.C., Fisher, G., Hsu, R.H., Lau, R., Bowen, B.P., Hromada, S., Northen, T., and Arkin, A.P. (2018). Deciphering microbial interactions in synthetic human gut microbiome communities. Mol. Syst. Biol. *14*, e8157.

Venugopal, V., and Nadkarni, G.B. (1977). Regulation of the arginine dihydrolase pathway in *Clostridium sporogenes*. J. Bacteriol. *131*, 693–695.

Walter, J., Maldonado-Gómez, M.X., and Martínez, I. (2018). To engraft or not to engraft: an ecological framework for gut microbiome modulation with live microbes. Curr. Opin. Biotechnol. *49*, 129–139.

Wick, R.R., Judd, L.M., Gorrie, C.L., and Holt, K.E. (2017). Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput. Biol. *13*, e1005595.

Widder, S., Allen, R.J., Pfeiffer, T., Curtis, T.P., Wiuf, C., Sloan, W.T., Cordero, O.X., Brown, S.P., Momeni, B., Shou, W., et al. (2016). Challenges in microbial ecology: building predictive understanding of community function and dynamics. ISME J *10*, 2557–2568.

Wildenauer, F.X., and Winter, J. (1986). Fermentation of isoleucine and arginine by pure and syntrophic cultures of *Clostridium sporogenes*. FEMS Microbiol. Lett. *38*, 373–379.

Wood, D.E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. Genome Biol *20*, 257.

Wu, M., McNulty, N.P., Rodionov, D.A., Khoroshkin, M.S., Griffin, N.W., Cheng, J., Latreille, P., Kerstetter, R.A., Terrapon, N., Henrissat, B., et al. (2015). *Genetic determinants of in vivo fitness and diet responsiveness in multiple human gut Bacteroides*. Science *350*, aac5992.

Wymore Brand, M., Wannemuehler, M.J., Phillips, G.J., Proctor, A., Overstreet, A.-M., Jergens, A.E., Orcutt, R.P., and Fox, J.G. (2015). The altered schaedler flora: continued applications of a defined murine microbial community. ILAR J *56*, 169–178.

Xavier, J.B. (2011). Social interaction in synthetic and natural microbial communities. Mol. Syst. Biol. *7*, 483.

Xu, M., Guo, L., Gu, S., Wang, O., Zhang, R., Fan, G., Xu, X., Deng, L., and Liu, X. (2019). TGS-GapCloser: fast and accurately passing through the Bermuda in large genome using error-prone third-generation long reads. Preprint at bioRxiv. https://doi.org/10.1101/831248.

Ze, X., Duncan, S.H., Louis, P., and Flint, H.J. (2012). *Ruminococcus bromii* is a keystone species for the degradation of resistant starch in the human colon. ISME J *6*, 1535–1543.