1 **TITLE**

2 Endospores and other lysis-resistant bacteria comprise a widely shared core community

3 within the human microbiota

4

5 **AUTHORS**

6 Sean M. Kearney[1,2,3], Sean M. Gibbons[1,2,3], Mathilde Poyet[1,2,3], Thomas Gurry[1,2,3],

7 Kevin Bullock[2], Jessica R. Allegretti[4,5], Clary B. Clish[2], Eric J. Alm[1,2,3,*]

8 [1] Department of Biological Engineering, Massachusetts Institute of Technology,

9 Cambridge, MA, U.S.A.

10 [2] The Broad Institute, Cambridge, MA, U.S.A.

11 [3] The Center for Microbiome Informatics and Therapeutics, Cambridge, MA, U.S.A.

12 [4] Division of Gastroenterology, Brigham and Women's Hospital, Boston, MA, USA

13 [5] Harvard Medical School, Boston, MA, USA

14 * Corresponding author: ejalm@mit.edu

15

ABSTRACT

Endospore-formers in the human microbiota are well adapted for host-to-host transmission, and an emerging consensus points to their role in determining health and disease states in the gut. The human gut, more than any other environment, encourages the maintenance of endospore formation, with nearly 60% of genera in the microbiome carrying genes attributed to this trait. However, there has been limited work on the ecological role of endospores and other stress-resistant cellular states in the human gut. Here, we applied a culture-independent protocol to enrich for endospores and other stress-resistant cells in human feces, and provide evidence that organisms found in resistant states are adapted for cross-host dissemination. Furthermore, organisms with resistant cellular states respond to shared environmental signals that likely mediate colonization of recently disturbed human guts.

INTRODUCTION

To date, there is limited work investigating the relevance of stress-resistant cellular states in the propagation, survival, and function of organisms in the mammalian gastrointestinal (GI) tract. The gut is the only known environment with such a considerable abundance of organisms that form endospores, considered the most stress-resistant of all cell-types[1]. Anaerobic, endospore-forming Firmicutes are numerically dominant members of the GI tract of most animal species[2,3]. Within this group of organisms, the presence of genes for endospore formation suggests that

growth in the GI tract favors the maintenance of this large gene repertoire[2]. The apparent utility of these genes is to allow organisms to enter metabolically dormant states that aid in survival and transmission to new hosts. Passage through the GI tract is likely to trigger sporulation[4,5], but the mechanisms by which this process occurs and the signals that induce sporulation here are mostly unknown, even for well-studied pathogens like *Clostridioides difficile*. Resistant cellular states like endospores appear to be adaptive in the mammalian gut environment. It is likely that other non-endospore-forming taxa have evolved analogous resistance strategies for passing between hosts.

Many endospore-forming organisms in the human gut are in the class Clostridia, the most well-studied of which are the pathogens *C. difficile* and *Clostridium perfringens*[6,7]. However, Clostridia also includes abundant organisms not known to form endospores, like *Faecalibacterium prausnitzii*[8] and *Roseburia intestinalis*[9]. For *C. difficile*, the role of sporulation is central to disease etiology[10], particularly in patients who experience recurrence. Sporulation and rising levels of antibiotic resistance allow *C. difficile* to persist in the face of antibiotic assault, ensuring that it remains in the environment to rapidly re-colonize its host.

Among Clostridia that do not cause disease, multiple strains of endospore-forming organisms have the capacity to induce T regulatory cells and associated anti-inflammatory cytokines[11,12] involved in sensitivity to, for example, peanut antigen[13]. These organisms have recently been shown to provide pathogen resistance in neonatal mice[14]. Similarly, endospore-forming commensals of the murine GI tract have a central

role in mediating the induction of a Th17-type T helper cell response[15–17]. Many Clostridia also produce butyrate as an end-product of metabolism, which can regulate how immune cells interact with gut commensal bacteria[18–22]. This group of organisms also boosts production of serotonin by enterochromaffin cells in the intestine, crucial for motility in the gut[23]. An ecological understanding of sporulation and induction of other resistant states could be informative for how these phenotypes interact with host immunity. For example, such an understanding could inform whether inflammation acts positively or negatively on endospore formation, or whether endospores themselves have immunomodulatory effects.

Environmental stress resistance protects cells faced with unfavorable conditions. The signals triggering resistance are likely quite varied. Even for well-studied endospore-forming bacteria, inducing sporulation *in vitro* can be difficult, and across strains of one species, signals that induce sporulation in one strain may be insufficient to induce sporulation in others[24]. Further, even organisms that abundantly form endospores in their native environment may not do so under conditions permitting vegetative growth. For instance, *Paenibacillus larvae* in honeybees will only form endospores *in vitro* under idiosyncratic conditions designed to mimic the host environment[25]. Similarly, certain strains of *Clostridium perfringens* rarely form endospores *in vitro* unless exposed to a specific set of environmental stressors[24]. The discrepancy in phenotype of organisms in their native environments compared to *in vitro* argues for culture-free approaches to investigate such phenotypes *in situ*. Enriching for stress-resistant cells in environmental samples provides a means to uncover the actual context in which these states form.

Here, we investigate which organisms are present as endospores or as other resistant cell types in the human gastrointestinal tract. We modified previously described methods to enrich fecal samples for endospores and obtain paired bulk community and resistant fraction 16S rDNA sequence data for 24 healthy individuals and one individual across 24 days. We consistently enriched for putatively endospore-forming taxa in all samples, as well as other taxa, predominantly from the Actinobacteria phylum, that show high levels of lysis resistance. We compared resistant OTUs (rOTUs) and non-resistant OTUs (nOTUs) to identify ecological characteristics differing between these groups. Using a database of rOTUs, we find consistent signals for these organisms in their responses to a variety of disturbances across multiple independent data sets. Overall, we show a tight association between the ecological role of these resistant organisms and their distribution within and across human hosts.

RESULTS AND DISCUSSION

Sequencing resistant fraction reveals resistant taxa present in human feces

We modified a culture-independent method[26] to generate paired bulk community and resistant fraction 16S rDNA amplicon data from human feces (Figure 1A). Aggregating the data across our cohort, we see expansion of classes with known endospore-formers in the resistant fraction: Clostridia, Erysipelotrichia, and Bacilli (Figure 1B). We also see

107 depletion of classes lacking endospore-formers (Bacteroides, Betaproteobacteria,
108 Verrucomicrobia, Gammaproteobacteria).
109
110 Organisms in the class Actinobacteria were enriched in the resistant fraction, but lack
111 genes considered essential for endospore formation. Although exospore formation is
112 well documented in some families of Actinobacteria (e.g. Actinomycetaceae and
113 Streptomycetaceae), these families have only modest representation in our data. We
114 see high-level resistance primarily from *Bifidobacterium* and *Collinsella*, whose
115 representative genomes lack orthologs for genes thought to be essential to exospore
116 formation.
117
118 We suspect that high level resistance in the Actinobacteria is mediated primarily by
119 resistance to lysozyme conferred by cell wall structures common to Actinobacteria[27].
120 Lysozyme is one of the most common and important defense mechanisms used by
121 neutrophils, monocytes, macrophages, and epithelial cells[28,29]. It is abundant in human
122 milk, a source of *Bifidobacterium* species transferred to breast-feeding infants, and in
123 saliva and mucus, where it serves an antibacterial role[30]. Attempts to deplete
124 Actinobacteria with achromopeptidase, which has previously been shown to break down
125 Actinobacterial peptidoglycan, had variable efficacy across samples (data not shown).
126 Thus, factors other than cell wall structure may contribute to Actinobacteria resistance.
127
128 To quantify the extent of lysis resistance, we calculated the proportion of normalized
129 reads for each 100% OTU (unique sequence) in the resistant fraction to the sum of its

reads in the bulk community and the resistant fraction. We then obtain a finite quantity even for organisms not observed in one of the paired samples. When the proportion exceeds 0.5 we call an OTU enriched in the resistant fraction (Figure 1A). An OTU is enriched in at least one of the samples in which it is present is considered a resistant OTU (rOTU), and non-resistant (nOTU) otherwise. Using the above definitions, all of the rOTUs are either Firmicutes or Actinobacteria (Figure 1C). In fact, when grouping OTUs at the genus level, the top two most enriched genera (*Bifidobacterium*, *Collinsella*) are both Actinobacteria.

Resistant fractions consist of a few dominant and many rare OTUs

In order to investigate ecological properties of the resistant cell fraction, we first examined the community structure of resistant fractions and compared these to their bulk community counterparts. Using alpha diversity metrics, it is clear that resistant fractions are significantly less diverse than their bulk community counterparts (Figure 2A). However, this difference is driven largely by differences in evenness between the communities, as enrichments are significantly less even. The OTU richness between resistant fractions and bulk community samples tend to be no different. Thus, in general, resistant fractions are dominated by a few highly abundant OTUs and many OTUs with just a single count.

rOTUs are more shared than nOTUs among individuals

Resistant fraction OTUs are shared between individuals more often than bulk community OTUs (Figure 2B). By comparing the ratio of OTUs found in a single person to OTUs found in multiple individuals, we find a significantly higher proportion of OTUs shared in the resistant fraction than in the bulk community (Fisher Exact Test, odds ratio = 1.68, p-value = 1.5e-45, Figure 2B). Furthermore, the majority of OTUs found in the resistant fractions are not observed in their bulk community counterparts. We interpret these results to mean that the resistant fractions comprise a seed bank of generally rare OTUs that are widespread within our cohort.

To further test the hypothesis that resistant states contribute to prevalence, we examined the frequency with which rOTUs were found among the bulk communities across individuals compared to nOTUs (Supplementary Figure 1). First, nOTUs, which are never enriched in the resistant fraction, are significantly less likely to be shared among multiple individuals than rOTUs (Mann Whitney U Test comparing the distribution of the number of individuals sharing each rOTU to the number of individuals sharing each nOTU, p-value = 4.8e-8). We again see this result by calculating the correlation between the frequency of resistance (the number of times an organism is enriched in the resistant fraction divided by the number of times it is observed) and sharedness (number of individuals an OTU is observed in divided by the total number of individuals), giving a weak, but positive and highly significant correlation (Spearman rho, correlation = 0.12, p-value = 7.3e-25; Kendall tau, correlation = 0.10, p-value = 1.97e-24). This result suggests that organisms that do not form resistant states are less likely to be found across multiple individuals than those that do. In fact, rOTUs tend to be less

dominant members of the community (median rOTUs = 1 count, median nOTUs = 2 counts, Mann-Whitney U Test, p-value = 3.5e-13). This result indicates that organisms found in resistant states generally have lower overall abundance. We suggest that organisms with resistant states sacrifice abundance within an individual for higher rates of dispersal across individuals.

Representation of organisms in resistant fractions is heterogeneous across and within individuals

A tradeoff between resistance and population size within a host suggests that there may be strong positive selection for such resistance phenotypes. Variation in this trait among related organisms could be indicative of selection. In order to visualize how much of a population is present in a resistant state within a given sample, we scaled 16S rDNA abundance data using V4 16S rDNA qPCR-based estimates of community size and defined the resistant fraction as the ratio of these scaled reads for each OTU. We plot this quantity on a phylogeny representing 99% OTUs (clustered at 99% nucleotide identity) present in at least 8 individuals and up to 24 individuals (Figure 3). First, we note the high variability in the resistant fraction within and across taxa (the average variation is over 50-fold within each taxon). For one Roseburia 99% OTU in particular, this quantity varies over 3 orders of magnitude, suggesting this OTU contains organisms present in a resistant state in some individuals, but not in others.

198 Furthermore, within a person, OTUs with the same genus classification can be
199 discordant in their degree of resistance. In the individual time series, for example, one
200 *Ruminococcus* 100% OTU is almost always enriched, and another is never enriched
201 (Supplementary Figure 4). The closest matching genomes to these two organisms show
202 differences in sporulation gene content, with the resistant *Ruminococcus* sharing 48/58
203 core sporulation genes[31], and the non-resistant only 41/58 (Supplementary Figure 5 and
204 Supplementary Table 5). We also see that spore maturation proteins *spmA* and *spmB*
205 vary in their presence in genomes of genera with variable enrichment phenotypes.
206 These genes are involved in spore cortex dehydration and heat resistance in *B. subtilis*
207 and *C. perfringens*, so their loss might contribute to differences in the recovery of
208 resistant cells in this work.
209
210 Formation of resistant cells itself might be selected on in this system. There is evidence
211 that the sporulation phenotype is evolving in mammalian guts, as several gut isolates of
212 *B. subitilis* lack genes that negatively regulate sporulation compared to their laboratory
213 counterparts[34]. Knowing which organisms can form resistant cells in a community does
214 not provide complete information about which organisms do. Formation of resistant
215 states *in vivo* seems to be highly context dependent. We also note that loss of a single
216 gene (i.e. *spo0A*) in *C. difficile* is sufficient for loss of sporulation, such that retaining
217 endospore formation requires strong purifying selection.
218
219 *rOTUs share signals for growth within an individual*
220

Previous evidence has shown that bile acids contribute to outgrowth of *C. difficile* endospores *in vivo*[35]. As a first pass to measure bile acid concentration in the gut, we tracked a *Bilophila* OTU in the time series. *Bilophila* are known to use taurine derived from taurocholic acid, a primary bile acid, as an electron acceptor for sulfite reduction[36]. Among rOTUs within the bulk community time series, the average Spearman correlation to *Bilophila* is 0.110 compared to 0.015 for nOTUs (Mann Whitney U Test p-value = 3.17e-14). The abundance distribution for *Bilophila* provides information about the abundance of rOTUs. As, taurocholate is a known germinant for several endospore-forming species[2], *Bilophila* abundance might act as a proxy for taurocholate concentration or the capacity for sulfite reduction, which requires reducing anaerobic conditions.

More broadly, the consistent correlations between rOTUs and *Bilophila* implied that rOTUs would in general be more correlated to each other. By hierarchically clustering the Spearman correlation profiles of all OTUs present for at least half of the time series (Figure 4A), we find that the rOTUs mostly fall into one of two clusters, with nOTUs in the other cluster (Fisher exact test, p-value = 3.13e-25). The average correlation between rOTUs in the time series to each other is 0.156 compared to 0.014 for nOTUs to each other (Wilcox rank sum test, p-value = 0): nearly all of the strong correlated behavior in this person is within this group of organisms (Figure 4A). We interpret this result to mean that the dynamic behavior of rOTUs is strongly coupled: these OTUs respond coherently to environmental signals, which are likely linked to bile acids given their correlations to *Bilophila*.

244

245 *rOTUs link growth to fatty acid metabolism*

246

247 To address whether bile-related signals relate to the dynamics of rOTUs in the time
248 series, we conducted untargeted metabolomics with standards for fatty acid
249 metabolism. We then calculated the Spearman correlations between the median
250 abundance profile of OTUs in the rOTU-dominated correlation cluster and metabolites
251 for which we had standard markers. This cluster tends to correlate positively with long-
252 chain saturated fatty acids, and negatively with long-chain polyunsaturated fatty acids
253 and, notably, taurocholate (Supplementary Table 3). We suspect that taurocholate
254 metabolism by members of this group drives down the concentration of taurocholate in
255 stool. Additionally, saturated fatty acid concentration in the stool measures fatty acids
256 escaping absorption in the small intestine. This process would be negatively impacted
257 by microbial metabolism of taurocholate, as it more efficiently emulsifies saturated fats
258 than glycine-conjugated primary bile acids[36]. Fecal concentrations of taurocholate
259 reflect secretion of unmetabolized taurocholate, which should increase if taurocholate
260 metabolism by the gut microbiota decreases.

261

262 *Resistant cells lose resistance in response to physiological bile acid concentrations*

263

264 As a more direct test of the coupling of rOTUs to bile acid concentration, we dosed
265 ethanol-treated feces (to kill vegetative cells without the additional harshness of the
266 resistant fraction DNA enrichment protocol) with increasing concentrations of bovine bile

in aqueous solution. We then measured the depletion of OTUs from the endospore-enrichment using 16S rDNA sequencing (Figure 4B). When correcting for biomass via qPCR, nearly 20% of OTUs identified in the resistant fraction apparently germinated in response to bile acids (log-link quasipoisson generalized linear model, p-value < 0.05, Supplementary Table 4). The true fraction of resistant cells that lose resistance in response to bile acids is likely higher, as many endospores require an activation step (i.e. heating at 80°C or treatment with lysozyme as for *C. difficile*[37]) before they will respond to germinants.

Notably, most ethanol-resistant OTUs began to show a germination-like response at 0.5% bile (Figure 4B), which is near the concentrations found in the human small intestine[,38]. Although Clostridia and other putative endospore-formers make up the majority of organisms that lose resistance in response to bile acids, genera in the Actinobacteria and other resistant cells also show this response when approaching physiological concentrations. These conserved responses suggest that the same cues can mediate loss of resistance in distantly related organisms, similar to the conserved resuscitation response of dormant bacteria to peptidoglycan[39].

rOTUs exhibit shared dynamics in diverse contexts

Correlated behavior, increased prevalence, and shared signals for growth among rOTUs indicated that these organisms might exhibit a global response during disturbances of various kinds. To test this hypothesis, we made a sequence database of

rOTUs within our cohort, and used this database to identify putative rOTUs in other datasets (Figure 5A). We expected that increased prevalence and shared signals for growth would lead to enhanced colonization of the developing infant gut microbiota[40]. The lysozyme-resistant members of the Actinobacteria and Bacillales dominate the infant gut microbiota for most of the first 80 days of life and do not equilibrate until the infant starts a full adult diet (Figure 5B). Early colonization by these rOTUs connects a resistant state to development of the infant gut microbiome. Here, lysozyme resistance might be essential for semi-selective transmission of *Bifidobacterium,* as human breast milk is rich in lysozyme, potentially lysing non-resistant cells[41]. Others have shown endospore-formers negatively associate with vertical transmission from mother to infant[42], but other environmentally resistant states as in the Actinobacteria may be important for vertical transmission.

Depletion of endospore-forming clades is common during infection with *C. difficile*. We predicted a strong signal for rOTUs in individuals infected with *C. difficile*, due to its sporulation requirement for transmission[10]. We find a significant depletion of rOTUs dependent on *C. difficile* infection status (Figure 5C), with a serial depletion of rOTUs from healthy to first time diagnosis to recurrent patients[43]. Because of this depletion in rOTUs, we expected that fecal microbiota transplant (FMT) might transfer relatively more rOTUs than other OTUs[44]. Indeed, among OTUs shared with donors, 90% of rOTUs increase in abundance following FMT, compared to 77% for the rest of the community (Fisher exact test, p-value = 0.008) (Figure 5D).

We suspected that rOTUs are a particularly malleable component of the microbiota. To test this hypothesis, we measured the turnover of rOTUs in the time series of an otherwise healthy male individual who was infected by Salmonella[47]. New rOTUs almost completely replaced old rOTUs following this perturbation. By contrast, fewer OTUs from the rest of the community were lost and gained. This result holds both when examining the number of OTUs replaced (Fisher exact test, p-value = 6e-12) as well as the change in abundance of these OTUs (Figure 5E). We see again that rOTUs exhibit coherent responses to changes in the gut environment, most pronounced in systems with dramatic perturbations. Colonization of newly vacant niches favors rOTUs, likely transmitted in an endospore or other resistant state to germinate in an environment replete with nutrients (including untransformed bile acids). In the absence of a fully functioning microbiota, rOTUs appear to fill open niches more readily than nOTUs.

CONCLUSION

Gut bacteria in the resistant fraction were more shared across individuals and showed more correlated dynamics compared to non-resistant organisms. Resistant taxa show greater turnover following large-scale disturbance events, as in the case of *C. difficile* and *Salmonella* infection, which suggests that many of these organisms are sensitive to environmental fluctuations and respond to stress by entering into a dormant, seed-like state. Environmental sensitivity and high turnover rates of resistant taxa provide an opportunity to manipulate the composition of the human gut microbiota through targeted perturbations and replacements. Because of the therapeutic relevance of Clostridia

336 endospores[11–14,23], determining the exact conditions that permit their replacement may
337 be of high value for future microbiota-based therapeutics. Here, we found that the
338 growth of many resistant organisms was associated with dietary fatty acids. If this result
339 extends to more individuals, one can imagine a therapeutic strategy coupling dietary
340 changes with introduced resistant cells to enable robust colonization and engraftment.

MATERIALS AND METHODS

See supplemental information for materials and methods

DATA AVAILABILITY STATEMENT

All amplicon sequencing data generated in this study have been can be accessed on the US National Center for Biotechnology Information SRA database under BioProject PRJNA389431. Metabolomics data and WGS metagenomic data will be made available on request.

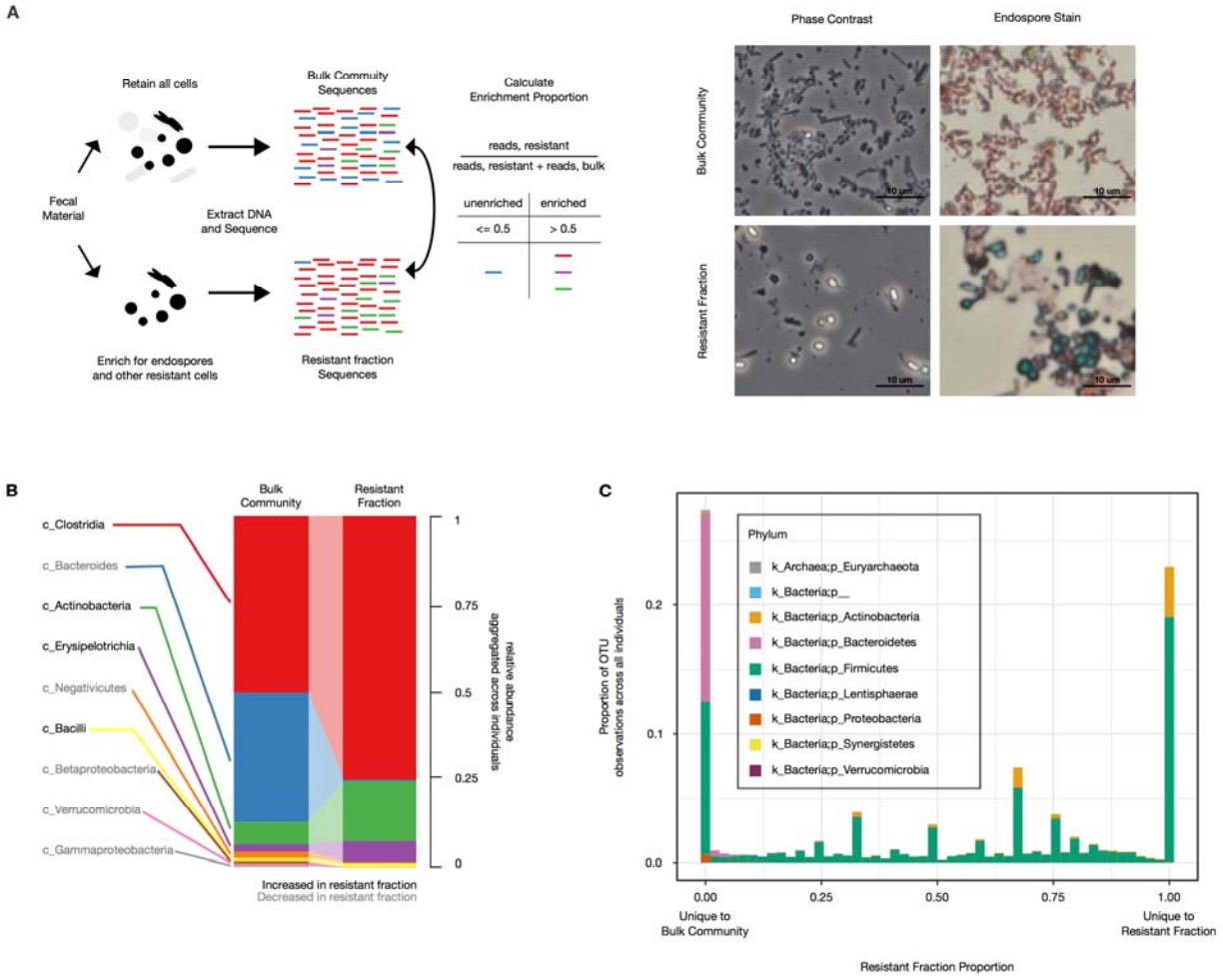
CODE AVAILABILITY STATEMENT

All custom scripts generated in R to analyze the data in this paper will be made available through GitHub (https://github.com/microbetrainer/Spores). The sequence processing pipeline is currently available online (http://amplicon-sequencing-pipeline.readthedocs.io/en/latest/)

Acknowledgments

We thank Fatima Hussain and Mathieu Groussin for extensive discussion and experimental advice. We thank the MIT BioMicro Center for sequencing service.

Funding was provided by the Broad Institute BN10 Training Grants. SM Kearney was funded by an NSF Graduate Research Fellowship.



364

365 **Figure 1. Resistant fraction sequencing of human fecal bacteria.** (A) Overview of
366 resistant cell enrichment and 16S rDNA sequencing protocol. Resistant fraction
367 samples are treated with a series of physical, enzymatic, and chemical lysis steps to
368 deplete vegetative cells. DNA from bulk community and resistant fraction samples are
369 extracted via a mechanical lysis protocol, and 16S rDNA libraries prepared.
370 Communities are analyzed to determine the change in abundance of each OTU in the
371 resistant fraction relative to the bulk community. (right) Phase contrast images of bulk
372 community and resistant fraction – phase bright cells are endospores. Endospores stain
373 green when heat fixed with malachite green, vegetative cells appear red from safranin
374 counter stain. (B) Representative results of 16S rDNA profile for bulk community and
375 endospore-enriched samples. Reads from each OTU are summed across 24 individuals
376 to give a meta-bulk and meta-endospore community. Phylogenetic classes in black text
377 increase with resistant fraction; gray text classes decrease with resistant fraction. (C)
378 Distribution of resistant fraction proportion across phyla aggregated across individuals
379 filtered to remove OTUs with single counts in a sample (for visualization purposes).
380 Colors represent phyla. OTUs with a resistant fraction proportion of 0 are absent from

381 the resistant fraction; OTUs with a resistant fraction proportion of 1 are absent from the
382 bulk community and only found in the resistant fraction.
383

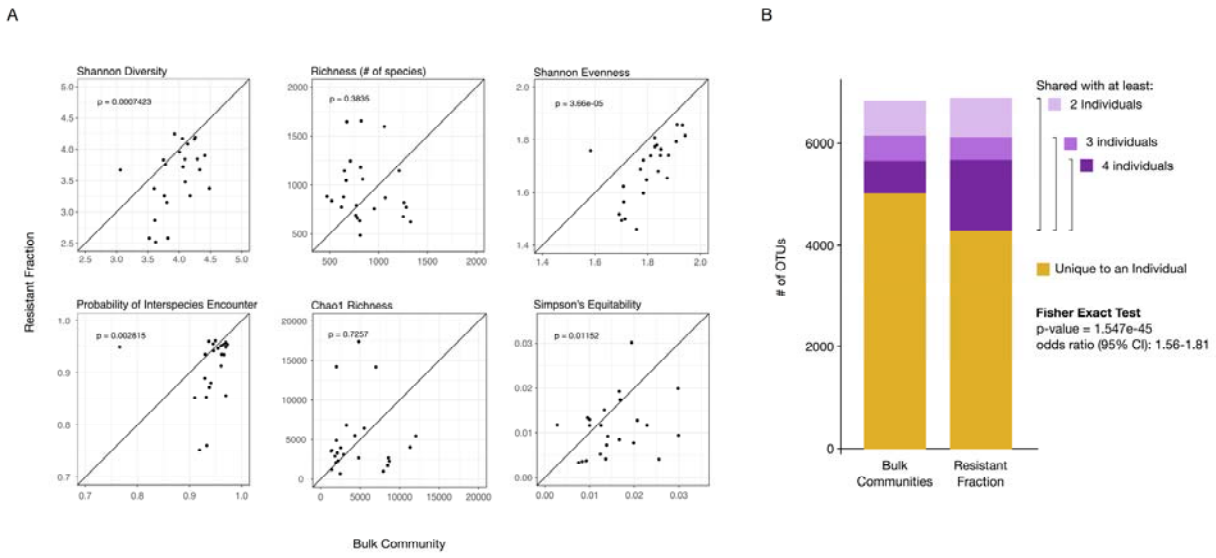


Figure 2. Resistant fraction OTUs are more shared across individuals than bulk community OTUs. (A) Alpha diversity metrics measured for the bulk community (x-axis) and resistant fraction enrichments (y-axis). P-values are for the test of differences between alpha-diversity metric distributions using paired Wilcoxon Rank Sum Test. (B) The number of OTUs found in only a single individual in the bulk communities and endospore enrichments are indicated in orange. The number of OTUs found in more than one individual are indicated through varying shades of purple. The odds ratio for the ratio of shared OTUs to unique OTUs for the endospore enrichments compared to the bulk communities is 1.69.

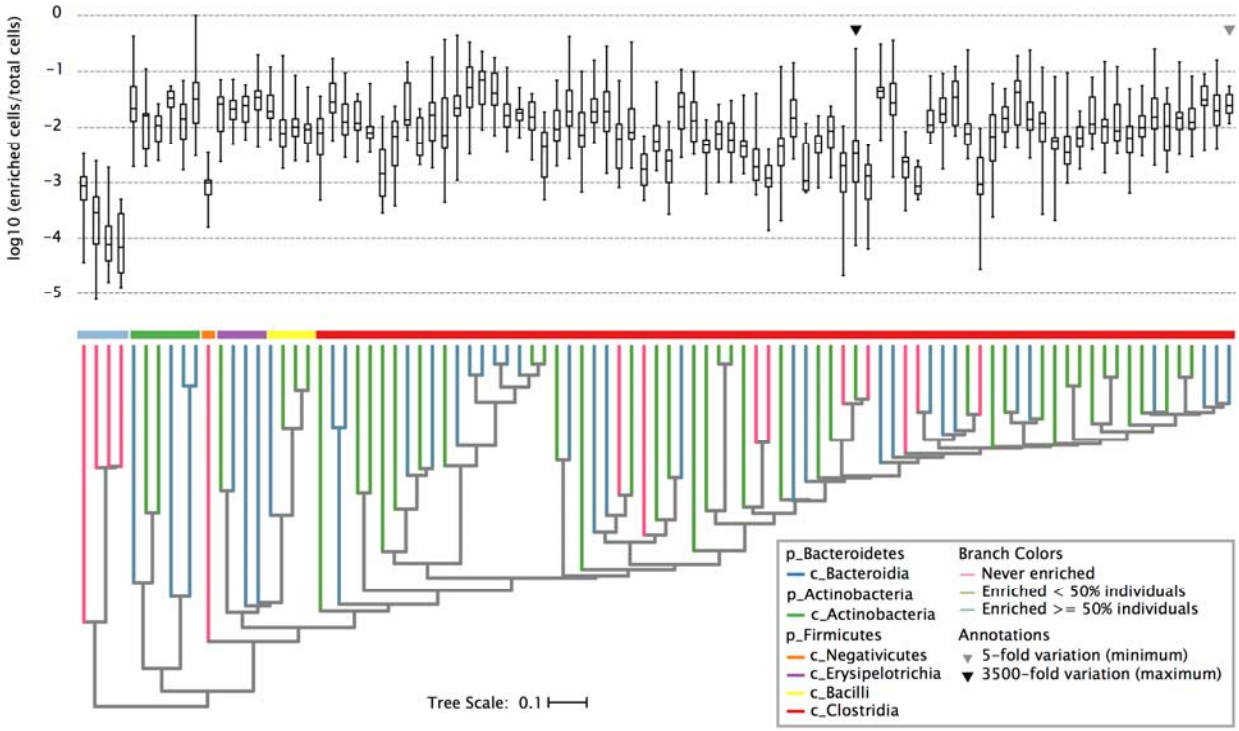


Figure 3. Taxa show heterogeneous patterns of resistant cell fractions across individuals. Phylogenetic placement of the fraction of resistant organisms for taxa present within at least 8 individuals estimated by the ratio of counts scaled by qPCR-estimates of biomass in the resistant fractions and bulk communities. Tree branch colors represent the degree to which a taxonomic group was enriched in the resistant fraction with pink branches never enriched and blue and green branches enriched at least once. Classes within each phylum are shown with a colored bar. Arrows indicate OTUs showing the maximum (black) and minimum (gray) within-OTU variability in enrichment scores.

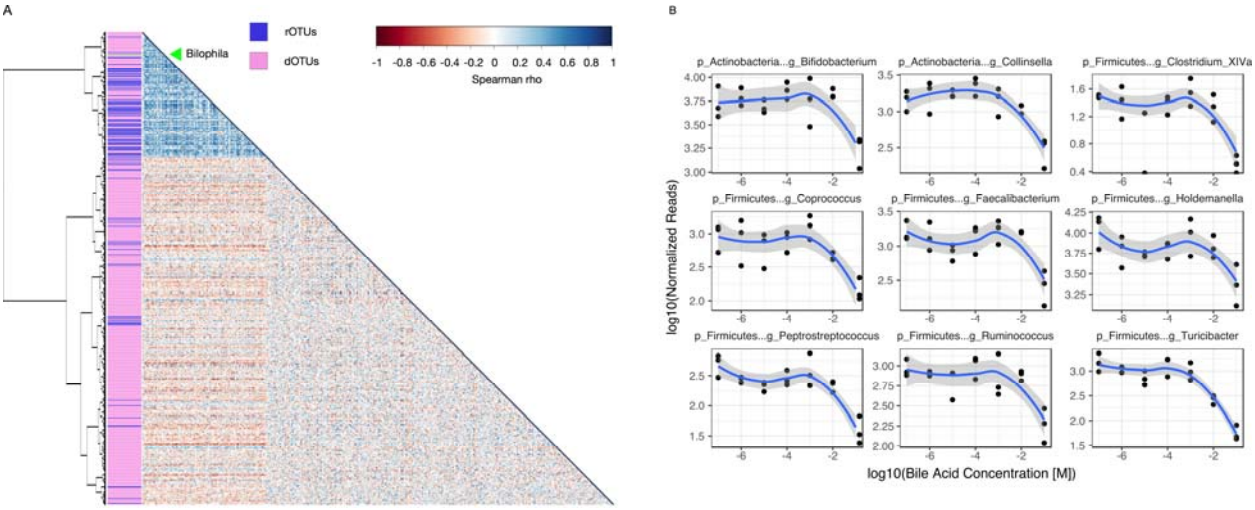


Figure 4. Common signals govern resistant state exit and growth in the GI tract. (A) Correlation matrix of OTU time series data within a single individual's bulk community hierarchically clustered (Ward's D) by Spearman correlation profile to other OTUs over time; bars at the tips of the Ward hierarchical cluster represent rOTUs (blue), nOTUs (pink), *Bilophila* (green). Strength of the correlation varies from red (perfectly anti-correlated) to blue (perfectly correlated). (B) Abundance of OTUs in the resistant fraction as a function of bile acid exposure for nine phylogenetically distant OTUs.

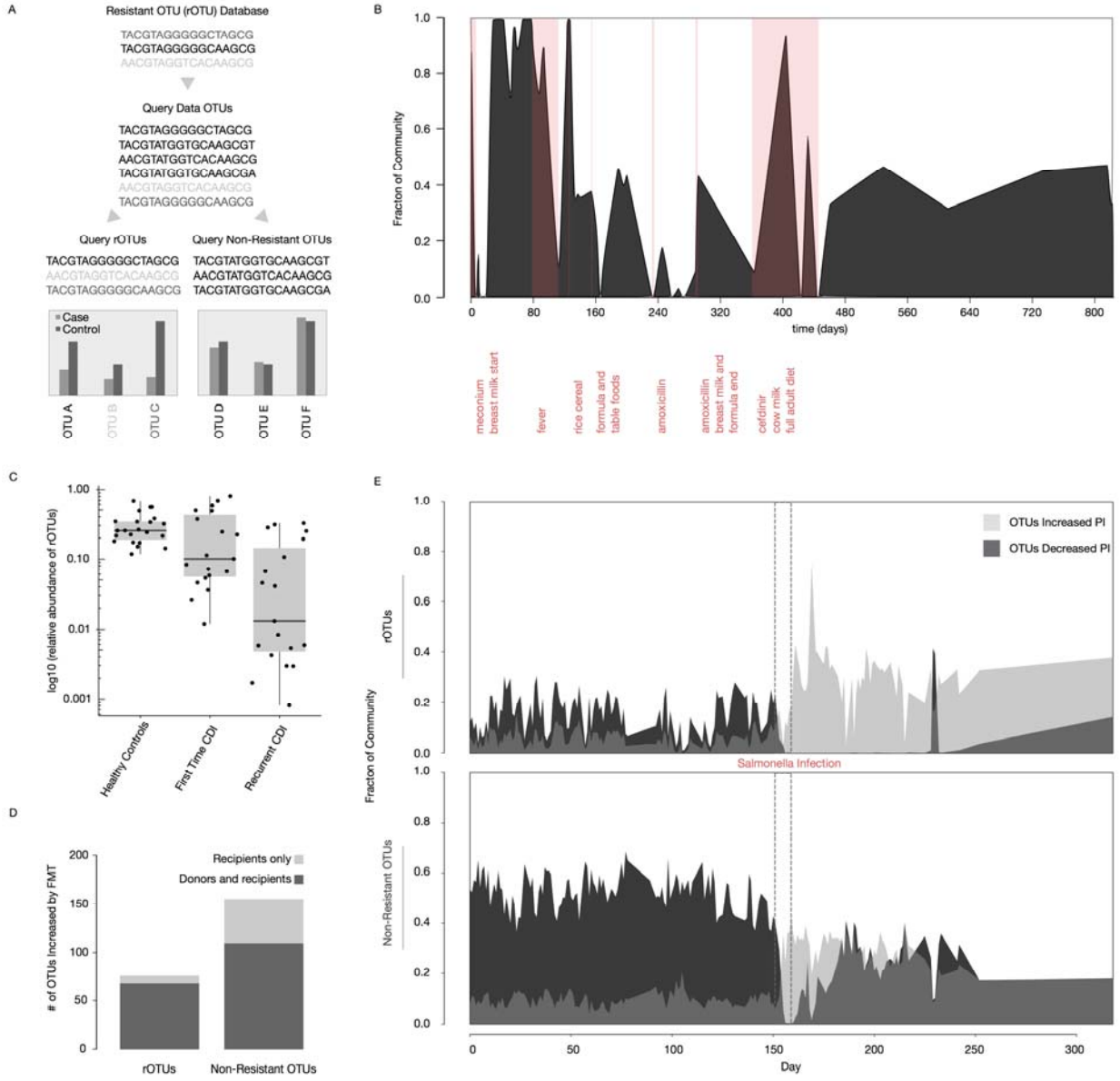


Figure 5. Resistant OTUs show disproportionate turnover in diverse contexts. (A) Overview of approach for identifying resistant-cell forming OTUs in 16S rDNA sequencing datasets. rOTU database sequences are matched to sequences in other datasets, and then patterns within those datasets among the identified rOTUs are determined. (B) Fraction of rOTUs present during microbial colonization of an infant gut annotated with major diet and health perturbations. rOTUs encompass both putative endospore-forming organisms and those not known to form endospores, but which possess a resistant state (Actinobacteria and non-endospore-forming Firmicutes) (C) Fraction of rOTUs present as a function of *C. difficile* infection status (fCDI = first time *C. difficile* diagnosis, rCDI = at least 3 episodes of *C. difficile* infection following initial treatment) (D) Fraction of rOTUs and all other OTUs (non-resistant OTUs) transferred from donors to recipients by fecal microbiota transplant. (E) Time series of rOTUs (top) and all other (non-resistant) OTUs (bottom) from a human male infected with

433 Salmonella, with OTUs significantly more abundant pre-infection (dark gray) and
434 significantly more abundant post-infection (light gray).
435

436

MATERIALS AND METHODS

Contact for reagent and resource sharing

Further information may be obtained from the Lead Contact Eric J. Alm (Email: ejalm@mit.edu; address: Massachussetts Institute of Technology Cambridge, MA, 02139, USA)

Experimental Model and Subject Details

Human Subjects

Human subject enrollment and sample collection was approved by the Institutional Review Board of the Massachusetts Institute of Technology (IRB Approval Number: 1510271631). Informed consent was obtained from all subjects. 12 male and 12 female healthy human subjects (age range 21-65) with no history of antibiotic use in the last six months were enrolled in the study. In total, 24 fecal samples were collected from these individuals and an additional 24 fecal samples were collected from one male individual (age 24) over 24 days for culturing and DNA isolation.

Bacterial Cell Culture Conditions

Identities and culture purity of bacterial species were confirmed by Sanger sequencing of the 16S rRNA genes followed by comparison to sequences in public databases. Bacteria were grown in their respective media for all experiments except where otherwise detailed. For routine culture, frozen glycerol stocks of each strain are first streaked onto agar supportive of growth before inoculation of single colonies into liquid medium. All bacteria were grown to log-phase at 37° C under anaerobic conditions (atmosphere 85% N_2, 5% H_2, 10% CO_2), except when preparing endospores.

Method Details

Fecal Sample Processing and Storage

Fecal samples were collected and processed in a biosafety cabinet within 30 minutes of defecation. Samples (5 g) were suspended in 20 mL of 1% sodium hexametaphosphate solution (a flocculant) in order to bring biomass into solution as described previously[26]. Fecal samples were bump vortexed with glass beads to homogenize, and centrifuged at 50 x g for 5 min at room temperature to sediment particulate matter and beads. Triplicate aliquots of 1 mL of the supernatant liquid were transferred into cryovials and stored at -80° C until processing. For the time series, samples were collected at approximately 24-hour intervals to reduce variation from uneven sampling frequency.

Resistant Fraction Enrichment from Fecal Samples

We modified a previously published method[26] for endospore sequencing to increase throughput and decrease signal from contaminating, non-endospore forming organisms. Fecal samples previously frozen at -80° C were thawed at 4° C prior to use, and 500 µL was aliquoted for resistant fraction, while the remaining 500 µL was saved for bulk community DNA extraction. Samples were centrifuged at 4° C and 10,000 x g for 5 minutes, washed and then resuspended in 1 mL Tris-EDTA pH 7.6. Samples were heated at 65° C for 30 minutes with shaking at 100 rpm and then cooled on ice for 5 minutes. Lysozyme (10 mg/mL) was added to a final concentration of 2 mg/mL and the samples were incubated at 37° C for 30 minutes with shaking at 100 rpm. At 30 minutes, 50 uL Proteinase K (>600 mAU/ml) (Qiagen) was added and the samples incubated for an additional 30 minutes at 37° C. Next, 200 uL 6% SDS, 0.3 N NaOH solution was added and the samples incubated for 1 hour at room temperature with shaking at 100 rpm. Samples were then centrifuged at 10,000 rpm for 30 minutes. At this step, a pellet containing resistant endospores should be visible or slightly visible in the sample, and the pellet is washed three times at 10,000 x g with 1 mL chilled sterile ddH2O. The pellet is then resuspended in 100 uL ddH2O, and treated with 2 uL DNAse I (Ambion) to remove residual contaminating DNA with incubation at 37° C for 30 min. The DNAse is killed by addition of 10 µL Proteinase K (Qiagen) and incubation at 50° C for 15 minutes, followed by incubation at 70° C for 10 minutes to inactivate Proteinase K. At this step, microscopic examination of samples is used to confirm the presence of phase-bright (or phase-dark) spores. The sample is then ready for downstream extraction and sequencing.

505 **Extraction of Nucleic Acids**

506

507 We extracted DNA from both the original sample suspended in sodium
508 hexametaphosphate and the output of the resistant fraction. Both the original sample
509 and the resistant fraction were extracted with MoBio PowerSoil Isolation Kit (MoBio
510 Laboratories, Inc.) with three 10 minute bead-beating steps followed by sequential
511 collection of ⅓ of the solution to enhance recovery of endospore DNA as shown
512 previously[46]. DNA was extracted from bacterial pure cultures, fecal enrichment cultures,
513 and endospores using the same protocol as for fecal samples in order to achieve
514 consistency between the methods. DNA from bacterial colonies for 16S rDNA Sanger
515 sequencing confirmation or qPCR was obtained by homogenizing colonies in alkaline
516 polyethylene glycol buffer as described previously [47].

517

518 **16S rDNA Library Preparation and Sequencing**

519

520 Libraries for paired-end Illumina sequencing were constructed using a two-step 16S
521 rRNA PCR amplicon approach as described previously with minor modifications [48]. The
522 first-step primers (PE16S_V4_U515_F, 5′ ACACG ACGCT CTTCC GATCT YRYRG
523 TGCCA GCMGC CGCGG TAA-3′; PE16S_V4_E786_R, 5′-CGGCA TTCCT GCTGA
524 ACCGC TCTTC CGATC TGGAC TACHV GGGTW TCTAA T 3′) contain primers U515F
525 and E786R targeting the V4 region of the 16S rRNA gene, as described previously [48].
526 Additionally, a complexity region in the forward primer (5′-YRYR-3′) was added to help
527 the image-processing software used to detect distinct clusters during Illumina next-

generation sequencing. A second-step priming site is also present in both the forward (5′-ACACG ACGCT CTTCC GATCT-3′) and reverse (5′-CGGCA TTCCT GCTGA ACCGC TCTTC CGATC T-3′) first-step primers. The second-step primers incorporate the Illumina adapter sequences and a 9-bp barcode for library recognition (PE-III-PCR-F, 5′-AATGA TACGG CGACC ACCGA GATCT ACACT CTTTC CCTAC ACGAC GCTCT TCCGA TCT 3′; PE-III-PCR-001-096, 5′-CAAGC AGAAG ACGGC ATACG AGATN NNNNN NNNCG GTCTC GGCAT TCCTG CTGAA CCGCT CTTCC GATCT 3′, where N indicates the presence of a unique barcode.

Real-time qPCR before the first-step PCR was done to ensure uniform amplification and avoid overcycling all templates. Both real-time and first-step PCRs were done similarly to the manufacturer's protocol for Phusion polymerase (New England BioLabs, Ipswich, MA). For qPCR, reactions were assembled into 20 µL reaction volumes containing the following: DNA-free H_2O, 8.9 µL, HF buffer, 4 µL, dNTPs 0.4 µL, PE16S_V4_U515_F (3 µM), 2 µL, PE16S_V4_E786_R (3 µM) 2 µL, BSA (20 mg/mL), 0.5 µL, EvaGreen (20X), 1 µL, Phusion, 0.2 µL, and template DNA, 1 µL. Reactions were cycled for 40 cycles with the following conditions: 98° C for 2 min (initial denaturation), 40 cycles of 98 C for 30 s (denaturation), 52° C for 30 s (annealing), and 72° C for 30s (extension). Samples were diluted based on qPCR amplification to the level of the most dilute sample, and amplified to the maximum number of cycles needed for PCR amplification of the most dilute sample. For first step PCR, reactions were scaled (EvaGreen dye excluded, water increased) and divided into three 25-µl replicate reactions during both first- and second-step cycling reactions and cleaned after the first-and second-step using Agencourt AMPure XP-PCR purification (Beckman Coulter, Brea, CA) according to manufacturer

instructions. Second-step PCR contained the following: DNA-free H_2O, 10.65 µL, HF buffer, 5 µL, dNTPs 0.5 µL, PE-III-PCR-F (3 µM), 3.3 µL, PE-III-PCR-XXX (3 µM) 3.3 µL, Phusion, 0.25 µL, and first-step PCR DNA, 2 µL. Reactions were cycled for 10 cycles with the following conditions: 98° C for 30 s (initial denaturation), 10 cycles of 98° C for 30 s (denaturation), 83° C for 30 s (annealing), and 72° C for 30s (extension). Following second-step clean-up, product quality was verified by DNA gel electrophoresis and sample DNA concentrations determined using Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fisher Scientific). The libraries were multiplexed together and sequenced using the paired-end with 250-bp paired end reads approach on the MiSeq Illumina sequencing machine at the BioMicro Center (Massachusetts Institute of Technology, Cambridge, MA).

qPCR

For testing of the resistant fraction protocol, qPCR was carried out as described in the **16S rDNA Library Preparation and Sequencing** section. Total bacterial abundance was quantified using the same primers. For quantification of Firmicutes and Actinobacteria, primer sequences were obtained from[49]. Primers were used at the same concentrations as 16S primers, and annealing temperatures were adjusted to the appropriate temperature for the corresponding primer pairs.

16S rDNA Sequence Data Processing and Quality Control

572 Paired-end reads were joined with PEAR[50] using default settings. Sequence data files
573 were processed using a custom amplicon sequencing pipeline. Sequences at least 252
574 bp long after quality filtering (quality trim Q25) were retained. After quality filtering, the
575 complexity region between the adapters and the primer along with the primer sequence,
576 was removed. All sequences not matching the first 15 bases of the primer were
577 removed. Sequences were trimmed to 252 base pairs and clustered at 100% identity
578 with UCLUST. OTUs were classified using RDP [51,52]. Sequences that had fewer than 10
579 reads across individuals were excluded from downstream analysis. The resulting count
580 tables were used as input for analysis within R.

581

582 **Identifying High Confidence Endospore-Forming & Resistant OTUs**

583

584 We developed a workflow for identifying organisms showing increased abundance in
585 the resistant fraction relative to the bulk community. We use as input an OTU table with
586 normalized counts obtained using cumulative sum scaling as implemented in the
587 package metagenomSeq in R, which improves estimation of true differences between
588 samples that would be lost in scaling to the total number of reads due to differences in
589 read depth[53]. Because of biases due to PCR amplification and sequencing and within
590 OTU-heterogeneity, we avoided making strong quantitative claims about the proportions
591 of spores formed by any given OTU. Instead, we examined the frequency with which an
592 OTU was enriched in the endospore-enriched fraction outside of Poisson noise within
593 individuals. More precisely, we used the counts of an OTU in the bulk community and in
594 the resistant fraction to estimate the Poisson variance – if the difference between counts

was within a pre-specified (we used 2 standard deviations) window of the estimated variance, we excluded calculations on such OTUs within a sample. Next, because there were several OTUs found in the resistant fraction that were absent from the bulk community, we excluded calculations involving these OTUs, which would fail to provide a finite estimate of the level of resistant fraction and would have apparently deflated prevalence in the bulk community samples. For the remaining OTUs, an enriched OTU was one that had more counts in the resistant fraction than in the bulk community and an unenriched OTU was one that had fewer counts in the resistant fraction than in the bulk community. The level of enrichment was calculated as the ratio of counts in the resistant fraction to the sum of counts in the bulk community and in the resistant fraction.

To compile a list of high-confidence resistant fraction-enriched organisms, we took a similar strategy as before, but also included OTUs which had 0 counts in the bulk community but non-zero counts in the resistant fraction. The OTUs increased in abundance in the resistant fraction compared to the bulk community in more than half of the samples present (excluding singletons) were included in this list.

Prevalence and abundance comparisons

To compare prevalence among OTUs in the resistant fraction and the bulk community, we counted the number of times an OTU had non-zero abundance in one of our subjects in either the resistant fraction or the bulk community. We used the Mann-

Whitney U test both to check for differences in read depth across the fractions and to compare the centers of these distributions. In order to compare prevalence of OTUs categorized as enriched to those categorized as unenriched (again excluding singletons), we used the Mann-Whitney U test on the prevalence estimated for these OTUs within the bulk community across the categories.

To determine the effect of abundance on prevalence estimations as above, we ran binomial logistic regression with log-transformed normalized counts within the bulk community as the predictor variable and endospore-enrichment status (rOTU or nOTU) as the binary response variable. A significantly positive slope indicates that endospore-enriched organisms are more likely than unenriched to derive from abundant organisms in the bulk community.

Genomic Spore Gene Content

Protein sequences in *Bacillus subtilis subtilis 168* from genes identified as shared among all spore-forming Bacilli and Clostridia[31] were downloaded from UniProt (http://www.uniprot.org/) to make a spore gene database. All genomes as of August 2016 from 9 genera of the Clostridia in containing OTUs that were both significantly enriched at times in the resistant fraction and significantly unenriched were downloaded from NCBI. A standard tblastn approach was used to identify homologues in the downloaded genomes with the corresponding genes in the spore gene database. After identifying presence/absence of spore genes, genome spore gene profiles were hierarchically clustered using UPGMA on the binary distance (Jaccard) matrix.

641

642 **Metabolite profiling**

643 Metabolites were measured using liquid chromatography tandem mass spectrometry
644 (LC-MS) method operated on a Nexera X2 U-HPLC (Shimadzu Scientific Instruments;
645 Marlborough, MA) coupled to a Q Exactive hybrid quadrupole orbitrap mass
646 spectrometer (Thermo Fisher Scientific; Waltham, MA) methods. Stool samples
647 (200mg/mL in 1% sodium hexametaphosphate) were homogenized using a TissueLyser
648 II (Qiagen). Stool homogenates (30 µL) were extracted using 90 µL of methanol
649 containing PGE2-d4 as an internal standard (Cayman Chemical Co.; Ann Arbor, MI)
650 and centrifuged (10 min, 10,000 x g, 4°C). The supernatants (2 µL) were injected onto a
651 150 x 2.1 mm ACQUITY UPLC BEH C18 column (Waters; Milford, MA). The column
652 was eluted isocratically at a flow rate: 450µL/min with 20% mobile phase A (0.1% formic
653 acid in water) for 3 minutes followed by a linear gradient to 100% mobile phase B
654 (acetonitrile with 0.1% formic acid) over 12 minutes. MS analyses were carried out
655 using electrospray ionization in the negative ion mode using full scan analysis
656 over *m/z* 70-850 at 70,000 resolution and 3 Hz data acquisition rate. Additional MS
657 settings were: ion spray voltage, -3.5 kV; capillary temperature, 320°C; probe heater
658 temperature, 300 °C; sheath gas, 45; auxiliary gas, 10; and S-lens RF level 60. Raw
659 data were processed using TraceFinder 3.3 (Thermo Fisher Scientific; Waltham, MA)
660 and Progenesis QI (Nonlinear Dynamics; Newcastle upon Tyne, UK) software for
661 detection and integration of LC-MS peaks.

662

663 **Bile germination tests**

Treatment of fecal samples with ethanol has previously been shown to allow culture-based recovery of endospore-forming organisms[2]. To this end, fresh fecal samples were homogenized in 50% ethanol (250 mg/mL), incubated for 1 hour under aerobic conditions with shaking at 100 rpm, and washed three times (5 min, 10,000 x g) with sterile water to remove residual ethanol. Serial dilutions from 1e-4-10% (w/v) bile bovine oxgall (Sigma) were prepared in sterile water and 2.5 mL ethanol-treated fecal suspension mixed in triplicate with 2.5 mL each of these bile solutions. Samples were incubated under aerobic conditions for 2 hours at 37° C with 200 rpm shaking, and then transferred to -80° C prior to resistant fraction extraction and 16S rDNA library preparation.

Bile germination analysis

We transformed 16S rDNA sequencing counts generated by the bile germination tests again using the cumulative sum-scaling transformation[53]. Under the assumption that cells in the resistant fraction can only decrease or remain the same during treatment, we searched for negative relationships between bile acid concentration and abundance that would indicate and OTU had germinated. To identify significant negative relationships, we first fit a generalized linear model (GLM) with a log-link quasi-Poisson distribution to the normalized counts of OTUs present in the control sample with bile acid concentration as the predictor variable. We then identified the OTU with the strongest positive trend in the data (that with the highest positive slope and lowest p-

687 value). We assume that OTUs increase due only to compositional effects (that is, this
688 OTU has not germinated but its abundance apparently increases due to loss of other
689 OTUs), and we use the slope estimated from the fit of this model to detrend the other
690 dose-response data so as to constrain the abundance of this apparently increasing OTU
691 to be constant. We do so by dividing counts of all OTUs by exp(slope*bile acid
692 concentration), which is also a measure of the depletion of the endospore-enrichment
693 biomass. From this detrended dose-response data, we again fit a quasipoisson GLM
694 and identify putatively germinating OTUs as those having a significant ($p < 0.05$)
695 negative slope.
696
697 **Analysis of Infant Gut Time Series**
698
699 SRA files containing 16S rDNA Sequences were downloaded from Genbank under
700 accession no. SRA012472)[40]. Sequences were generated using a Roche 454
701 pyrosequencer. In order to simplify analysis of the dataset, these sequences were again
702 processed using the protocol outlined for processing of the original dataset in this paper.
703 However, sequences were quality trimmed using Q20 to 230 base pairs, and the
704 retained sequences were used to call 100% OTUs. OTUs were assigned taxonomies
705 using RDP and 100% OTUs were collapsed into taxonomic names. As very few
706 sequences matched between datasets when using uclust, these taxonomic names were
707 instead used to identify organisms as potential resistant cell-formers based on the
708 correspondence to the RDP-assigned taxonomic names of high confidence resistant
709 cell-formers identified previously. While this approach loses information given the noted

heterogeneity in resistance phenotypes even among closely related strains, the original sequences themselves are still proxies for having this phenotype, and so the results of such analysis must be interpreted keeping this observation in mind.

The relative abundance of organisms identified in the infant gut time series as putative resistant-cell formers were summed, and the dynamics of this resistant cell-forming population in the infant gut was visualized over time.

Analysis of 16S rDNA sequence files from first time and recurrent *C. difficile* infection

The open reference 97% OTU table including RDP taxonomic annotations from Allegretti et al 2016 was used for this analysis[43]. OTU IDs were mapped using uclust to the corresponding genus level OTUs identified as rOTUs from this study. Patients were grouped either as healthy, first-time *C. difficile* infection (fCDI), or recurrent *C. diffiicle infection* (rCDI), and the fraction of rOTUs was calculated by summing their relative abundances within each patient. A Mann Whitney U test was used to determine whether there were significant differences in the total relative abundance of rOTUs across groups with a Bonferroni multiple hypothesis test correction.

Analysis of 16S rDNA sequence files from fecal microbiota transplant in relapsing *C. difficile* infection

This dataset was obtained from Youngster et al, 2014[44]. To simplify analysis, an existing closed-reference GreenGenes 97% OTU table generated by the original authors was used. Closed-reference OTU IDs were mapped back to GreenGenes reference sequences, and sequences were assigned to the resistant cell-former database sequences again using uclust as for the adult time series.

Unique pre-FMT, post-FMT, and donor samples were separated in the dataset. We again identified organisms that had significantly different relative abundance (Benjamini-Hochberg adjusted Mann-Whitney U test $p < 0.05$) across the groups for our analysis. We again obtained four categories of OTUs: nonresistant and resistant cell-formers that were elevated in the donor and the post-FMT samples relative to the pre-FMT samples. We used the Fisher exact test on the contingency table containing the number of OTUs in each of the previously mentioned categories to identify whether OTU engraftment from the donor was different across the groupings.

Analysis of 16S rDNA sequence files in adult time series pre- and post-Salmonella Infection

Illumina HiSeq sequencing files containing 16S rDNA sequences from the stool of a healthy adult male[45] were downloaded and processed as described for the original dataset in this paper, except that sequences were trimmed to 101 base pairs as described previously before calling 100% OTUs due to the use of shorter read sequencing technology. Sequences were assigned to the resistant cell-former database

sequences using uclust constrained with the parameters: --id 99 –usersort –libonly, in order that sequences from this dataset would be assigned only to resistant cell-formers.

In order to assess the presence of differential turnover between resistant and non-resistant cell formers in this dataset, we identified organisms that had significantly different relative abundance (Benjamini-Hochberg adjusted Mann-Whitney U test $p < 0.05$) before Salmonella infection starting at day 151 (days 0-150) and after the end of infection at day 159 (days 160-252). We partitioned these OTUs into four sets for our analysis: non-resistant and resistant cell formers whose median abundance was higher post-infection and those whose median abundance was lower post-infection. We used the Fisher exact test on the contingency table containing the number of OTUs in each of the previously mentioned categories to identify whether the OTU turnover was different across the groupings.

Supplementary Methods

Preparation of Endospores of *Bacillus subtilis*

Endospores of *B. subtilis* for testing resistant fraction protocols were prepared as described previously[54]. *B. subtilis subtilis* strain AG174 (kindly provided by Alan Grossman, MIT) was streaked onto nutrient agar and grown overnight. A single colony was inoculated into 25 mL Difco Sporulation Medium (DSM), and grown at 37° C until mid-log phase (OD600 = 0.5), and diluted 1:10 into 250 mL of pre-warmed 37° C DSM

in a 2 L flask. Cells were grown for 48 hrs at 37° C and 150 rpm under aerobic conditions. When 90% of cells were free phase-bright spores, cultures were centrifuged at 10,000 x g and washed twice with 200 mL 4 C sterile distilled water. The pellet was resuspended in 200 mL cold distilled water and left at 4° C overnight, before repeating the washing procedure. No further purification steps were used as only spores were observed with microscopic examination. Spore viability and counts were confirmed by germination and growth after 24 hours on nutrient agar at 37° C.

Fluorometric Quantification of Dipicolinic Acid Content

Fecal samples were thawed at 4° C and 10 mg (50 µL) suspended in 0.9 ml sodium acetate buffer and 0.1 ml aluminium chloride ($AlCl_3$, 0.5 M) as described previously[50]. Samples were microwaved in microcentrifuge tubes for 20 s to disrupt endospores. After cooling on ice for 10 minutes, samples were centrifuged at 10,000 x g for 10 min, and 100 µL of the sample supernatant mixed with 100 µL terbium chloride solution ($TbCl_3$, 30 µM) in clear-bottom black 96-well microtiter plates. Fluorescence was immediately measured using a plate reader (Synergy H1) with the following settings: time-resolved fluorescence (delay 50 µs, interval 1200 µs) at an excitation wavelength of 272 nm, emission wavelength of 545 nm, and 10 endpoint readings per sample at 30°C. The number of spores in the samples was determined using standard addition method with endospores of *B. subtilis* or dipicolinic acid. Endospore content was expressed as equivalents of *B. subtilis*.

802

803 **Fecal spore-former culture library & isolate revival**

804

805 Thousands of isolates were collected from 12 healthy donor fecal samples (manuscript
806 in preparation), and isolates from a single donor from the cohort in this study was used
807 for all isolate experiments. Fecal material was diluted 1:15 in anaerobic PBS + 0.1% L-
808 cysteine and homogenized in an anaerobic chamber (Coy Anaerobic Systems, 20%
809 CO_2; 5% H_2; 75% N_2). 500 ul aliquots of the homogenized fecal material were stored in
810 50% glycerol at -80 C for continued culturing. Samples were treated with equivolume
811 amounts of 100% ethanol for 1 hour, and serial dilutions were plated on Gifu Anaerobic
812 Medium (GAM) or Brain Heart Infusion, Supplemented (BHIS) (citation). Individual
813 colonies were selected after 24 hours of growth under anaerobic conditions at 37° C,
814 and re-streaked onto fresh agar plates. Single colonies were grown in liquid medium,
815 and archived in 50% glycerol at -80° C. Aliquots of liquid culture were taken for 16S
816 rDNA Sanger Sequencing by first amplifying with the 27F (5'-
817 AGAGTTTGATCMTGGCTCAG-3') and 1492R (5'-TACGGYTACCTTGTTACGACTT-3')
818 primers, and then submitted with the U515F (5'-GTGCCAGCMGCCGCGGTAA-3')
819 internal primer. Sequences were assigned taxonomy using the RDP database [51,52].

820

821 **Testing Resistance of Actinobacteria**

822

823 For revival, frozen glycerol stocks of bacteria were streaked in an anaerobic chamber
824 onto pre-reduced (24 h) BHIS or GAM agar medium and incubated overnight under

anaerobic conditions at 37° C. For testing Actinobacteria isolates for lysis resistance, plates with revived organisms were incubated for 4 weeks under anaerobic conditions at room temperature to induce nutrient starvation. Plates were scraped of colonies and colonies resuspended in 1% sodium hexametaphosphate and subjected to spore enrichment as described previously. Results from these experiments in Supplementary Figure 3A.

SUPPLEMENTARY RESULTS AND DISCUSSION

Quantification of fecal-resident endospores

Although endospore-forming organisms are dominant in the gastrointestinal tract, the abundance of endospores has not been well quantified. Using a unique marker of endospores, dipicolinic acid, others have estimated a range in human fecal samples between 1e06 and 1e09 endospores per gram wet weight of stool[56], with a geometric mean of 1e08. With the same protocol, we estimate the endospore burden at 1.5e07-4.0e09 (95% CI) endospores per gram, which represents 0.03-0.7% (95% CI) endospores per cell estimated by qPCR (Supplementary Figure 2).

Using modifications on previously described methods[26], we depleted vegetative cells in human fecal samples to enrich for resistant cells (Fig 1A). By qPCR on this resistant fraction, our estimate of the resistant cell fraction is slightly higher than that given by dipicolinic acid quantification: 2e08-7e09 (95% CI) resistant cells per gram

(Supplementary Figure 2A), giving 0.2%-27% resistant cells per cell (Supplementary Figure 2B). Culture-based studies of germinable endospores widen these estimates and generally deplete them finding 1e05-1e09 germinable endospores per gram of stool, giving a wide range of 0.3-140% endospores per cell. We assume that most fecal samples contain at least 1e07 (and at most 1e12) endospores, which is roughly 100 times the infectious dose for endospores of *C. difficile*.

Most Clostridial OTUs in the resistant fraction show genetic capacity for sporulation

Many prominent gut commensals, including members of the genera *Faecalibacterium* apparently form endospores *in vivo* and have most of the genes to do so (Supplementary Table 1, 5, and Supplementary Figure 6), but have not been shown to form endospores *in vitro*[57]. We believe that for *Faecalibacterium* species as for endospores of *P. larvae* and *C. perfringens*, *in vivo* conditions drive endospore formation, and these conditions are not readily replicated *in vitro*. Despite being an extensively studied, numerically abundant, prevalent, and disease-relevant commensal, endospore formation has not been demonstrated for this organism.

Clostridial OTUs in general are more shared, whether present as rOTUs or nOTUs

When restricting our analysis of prevalence to only the Clostridia, there is no difference in sharing between Clostridia rOTUs and nOTUs (Mann Whitney U Test, p-value = 0.9952). The majority of the sharing signal is phylogenetic – when comparing

prevalence of non-Clostridia nOTUs to Clostridia nOTUs and rOTUs combined, the difference in prevalence is very pronounced (Mann Whitney U Test, p-value = 3.0e-24). Even Clostridia OTUs not identified as rOTUs may still contain organisms with the potential to enter resistant states, but may not necessarily be doing so in all individuals or at all times. Alternatively, other conserved phylogenetic traits within the Clostridia could contribute to the widespread sharing of these organisms, but because sporulation is required for transmission of *C. difficile*[10], it seems possible that resistance is more broadly required for efficient transmission of the Clostridia in general.

When assessing differential sharing of rOTUs and nOTUs determined from the time series within the cross-section, we again find that the sharing is primarily associated with the Clostridia, as there is no significant difference in sharing among time series-defined Clostridia nOTUs and rOTUs (Mann Whitney U Test p-value = 0.10). In part, this result arises because there is a high degree of overlap between the OTUs defined as rOTUs and nOTUs in the time series and in the cross-section. Although infrequent, there are 27 inconsistent classifications (rOTUs become nOTUs and vice versa) and 268 consistent classifications between the time series and the cross-sectional data. Of the inconsistencies, all are Firmicutes, highlighting that endospore formation is likely driven by host- or strain-specific variability. However, it is clear that organisms linked to endospore formation are more prevalent across unrelated hosts.

If we compare the abundance of rOTUs and nOTUs from the Clostridia, we still find that 11/24 individuals have Clostridia rOTUs at a significantly lower abundance than nOTUs

(none have significantly higher rOTU than nOTU abundance). In general, when combining the independent tests of differences between the abundance of Clostridia nOTUs and rOTUs, we find a significant tendency for Clostridia rOTUs to be at lower abundance (Fisher's Method, $X^2_{df=48}$ = 289.6 p-value = 0, 0, all OTUs, Clostridia OTUs).

We wanted to determine if Clostridia drove a differential persistence signal in the time series. Comparing the persistence of only Clostridia nOTUs and rOTUs, we find that Clostridia nOTUs are significantly more persistent than rOTUs defined within the time series (Wilcoxon rank sum test, p-value = 2.5e-05), and for those Clostridia nOTUs and rOTUs shared with the cross-sectional data (Wilcoxon rank sum test, p-value = 2.4e-5). This result suggests a tradeoff in forming resistant cells and within-host persistence in this person. Only those Clostridia that are enriched in the resistant state have reduced within-host persistence. Similarly, the Actinobacteria (all of which are rOTUs) are also more ephemeral than nOTUs (p-value = 2.2e-5), suggesting a shared life history strategy in their resistance phenotype.

Organisms in Clostridium cluster XI (including *C. sordelli*, *C. scindens,* and *C.difficile*) link their growth with the concentration of bile acids in the colon. In the phylogenetically diverse group of organisms encompassed by Clostridium cluster IV (including the genera *Clostridium, Eubacterium, Ruminococcus, Coprococcus, Dorea, Lachnospira, Roseburia* and *Butyrivibrio*), many are linked to the production of secondary bile acids, and several also exhibit preferential growth in the mucus layer[20,58]. The enrichment for some of these organisms among rOTUs may then derive from initiation of vegetative

917 growth in the small intestine through bile acid-mediate germination, growth in the
918 mucosa, limited growth in the lumen, and enhanced shedding of endospores into the
919 lumen relative to vegetative cells – driving the observation of reduced abundance and
920 persistence. These pieces of evidence suggested that the dynamics of rOTUs may be in
921 part governed by bile acids and shared niche space, and as such, we expect these
922 organisms to show similar dynamic behavior in other contexts.
923
924 *Differences in abundance drive differences in persistence*
925
926 The difference in persistence between rOTUs and nOTUs is largely attributable to a
927 difference in relative abundance between these groups of organisms, with rOTUs
928 having significantly lower median abundance than nOTUs in the time series (Mann
929 Whitney U test, p-value = 3.8e-5). We recall that rOTUs across people tended to have
930 lower median abundance than nOTUs. By comparing the abundance distributions of
931 rOTUs and nOTUs within individuals, we find that this result holds for 16/24 individuals
932 sampled (Mann Whitney U Test, $p < 0.05$ for each test). Combining the results from
933 these tests (Fisher's Method, $X^2_{df=48} = 412.0$, p-value = 0) suggests that overall,
934 resistant OTUs are less abundant than non-resistant OTUs. We propose that the
935 reduced abundance of rOTUs accompanies reduced persistence within a host
936 associated with lysis resistance.
937
938 *Common microbiota is more likely to source from rOTUs than nOTUs*
939

Comparing the fractions of rOTUs and nOTUs present in each individual to assess sharing still reveals a tendency for increased sharing across individuals (Wilcoxon rank sum test, p-value = 4.3e-03, Supplementary Figure 3), meaning that individuals are more likely to source their microbiota from rOTUs than nOTUs. Similarly, comparing the frequency of appearance and disappearance of rOTUs and nOTUs in the time series gives the same observation of reduced persistence (Mann Whitney U Test, p-value = 2.7e-13). Together, these results imply that resistance states enable increased prevalence at the cost of within-host abundance. Because of the heterogeneity in endospore formation within species groupings, this result weakens when grouping OTUs at the 99% level (100% OTUs used previously), but is still significant (p-value = 8.6e-29, mean ρ = 0.083 (rOTUs), 0.047 (nOTUs)).

Metabolomic correlations to Bilophila

While there were no significant correlations between the *Bilophila* OTU and any individual bile acid in feces measured with untargeted metabolomics, it had a significant positive correlation (unadjusted p-value = 2.1e-06, Spearman ρ = 0.84) to a lipid marker of milk fat consumption, pentadecanoic acid[59]. Milk fat consumption, and not consumption of other kinds of fat, has previously been shown to lead to expansions of *Bilophila* in mouse models, which were correlated with increased taurocholate concentration in the gall bladder[36]. Thus, this particular OTU-metabolite relationship serves as an independent verification that experimentally validated relationships play out *in vivo*.

963

964 *Introduced endospore-formers compete with endogenous endospore-formers*

965

966 To further understand the dynamics of the endospore-forming fraction of the microbiota,
967 the subject of the time series began taking tablets once daily (3e5 CFU/tablet) of a
968 probiotic prepared as endospores (Miyarisan *Clostridium butyricum* Tablets) on day 6 of
969 the time course. On day 8, an OTU corresponding to *C. butyricum* became detectable in
970 both the bulk community and in the endospore fraction. These observations lead to two
971 conclusions: (1) the endospores are capable of outgrowth *in vivo* in the presence of a
972 complex community, and (2) the organisms are forming endospores *in vivo*, as the
973 levels present in stool exceed the dose in the tablet. *C. butyricum* administered as
974 endospores has previously been shown to prevent antibiotic-associated diarrhea in
975 children (Seki et al., 2003), as well as to ameliorate DSS-induced colitis in mice through
976 induction of IL-10 producing macrophages[60], suggesting that it grows *in vivo*, and is
977 sufficient to block colonization by pathogenic endospore-formers such as *C. difficile*.
978 Given these properties, we wanted to examine the dynamic behavior of this organism *in*
979 *vivo*, and relate it to the activity of endogenous endospore-formers.

980

981 When examining relationships of introduced *C. butyricum* to metabolites in the stool, we
982 found a strong anti-correlation to pentadecanoic acid (unadjusted p-value = 6.3e-05, ρ =
983 -0.72). Recall that *Bilophila* strongly correlated with pentadecanoic acid, and that most
984 other endospore-formers fell into the correlation cluster containing *Bilophila*. By
985 contrast, this OTU, which had been introduced as an endospore, exhibited the opposite

behavior. We believe that this result may be a signal of competitive exclusion: because endogenous endospore-forming commensals occupy the dominant niche, tethering their growth to shared signals, the introduced endospore-formers are restricted to a smaller, potentially orthogonal niche. Further work remains to determine the signals that mediate colonization of introduced endospores.

EXTENDED DISCUSSION

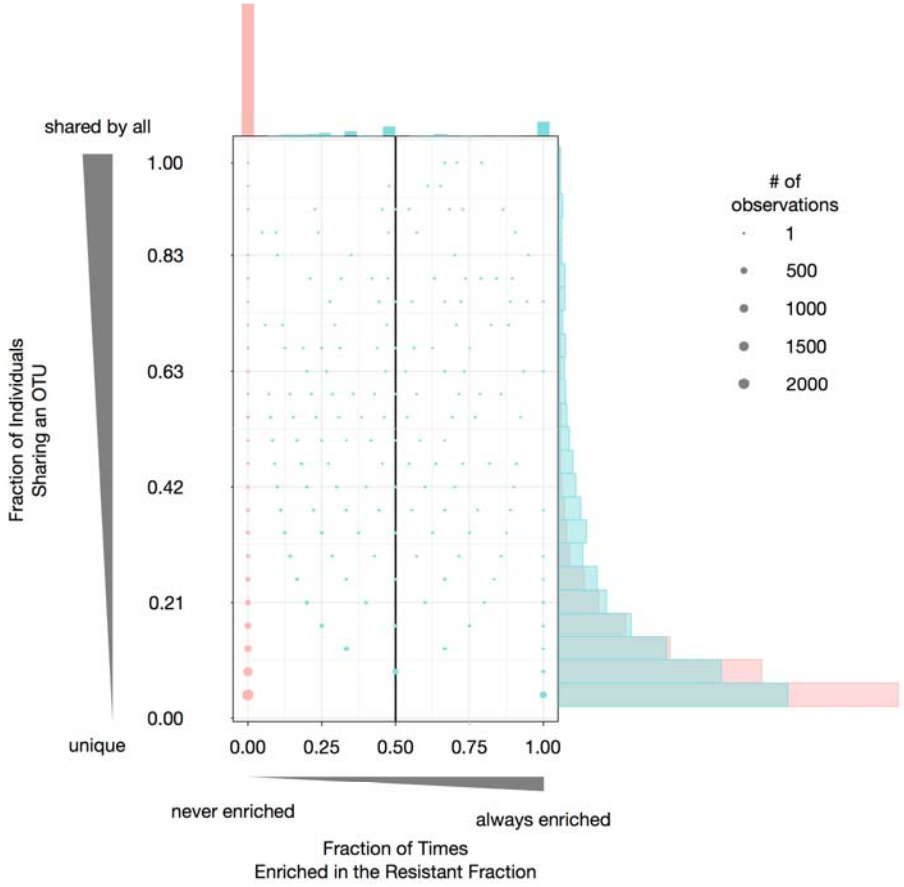
Anaerobic endospore-forming organisms dominate the mammalian gastrointestinal tract as no other studied environment. Why does this environment, compared to most others, favor the growth of organisms with a metabolically costly and risky strategy? While it seems obvious that this trait should increase transmission, and here we provide evidence for this effect through our observation of increased prevalence, it is not obvious that increased transmissibility is cost-free. We propose that a tradeoff exists in maintaining a sporulation phenotype: endospore formation incurs diminished vegetative growth, resulting in lower population sizes and thus decreased persistence within a host due to a greater probability of extinction within. Because sporulation and presumably other resistance traits seem to permit long-term survival outside of a host, this stability provides a means for maintaining these resistant-cell formers within host populations over time. By contrast, other strict anaerobic organisms in the gastrointestinal tract by definition exhibit the opposite strategy: instead of investing in endospore formation, they invest in larger population sizes that help maintain themselves within a given host.

If other organisms adapted to a gut environment did not have a strategy for transmission across hosts, we would only see endospore-formers across people. Of course, this is not the case, so how is it that these organisms transmit? There are a number of possibilities, but persister states may allow non-endospore-forming organisms to enter a metabolically inert state upon exit from the gastrointestinal tract. Toxin-antitoxin systems, associated with persistence in *E. coli*, are overrepresented in Bacteroidetes, Alpha- and Gammaproteobacteria[61], and Bacteroidetes are among the most metabolically inactive cells in human fecal samples[62]. However, we also observe that among the non-Clostridia nOTUs, abundance is higher within hosts, suggesting that simply by maintaining high population levels, these organisms increase their ability to transmit to new hosts. When host populations are high density, gut residents do not require long-term environmental survival; the probability of encountering a new host will be high. However, when host populations are low density, long-term environmental survival becomes a necessary strategy. Over longer evolutionary timescales, endospore-formers should be selected for in hosts with low population densities.

Resistant cells appeared to be primed to enter uncolonized gastrointestinal tracts. There may be a few reasons for this observation. Environmental resistance enables them to survive for long periods of time without diminished viability. Specific molecular indicators of a suitable host, such as bile acids, may permit germination and subsequent outgrowth under the right host conditions. Outgrowth seems to require strict anaerobic conditions and low reduction potentials as evidenced by the entrainment of rOTUs to the dynamic behavior of the strictly anaerobic sulfite-reducer, *Bilophila*. Similarly, growth

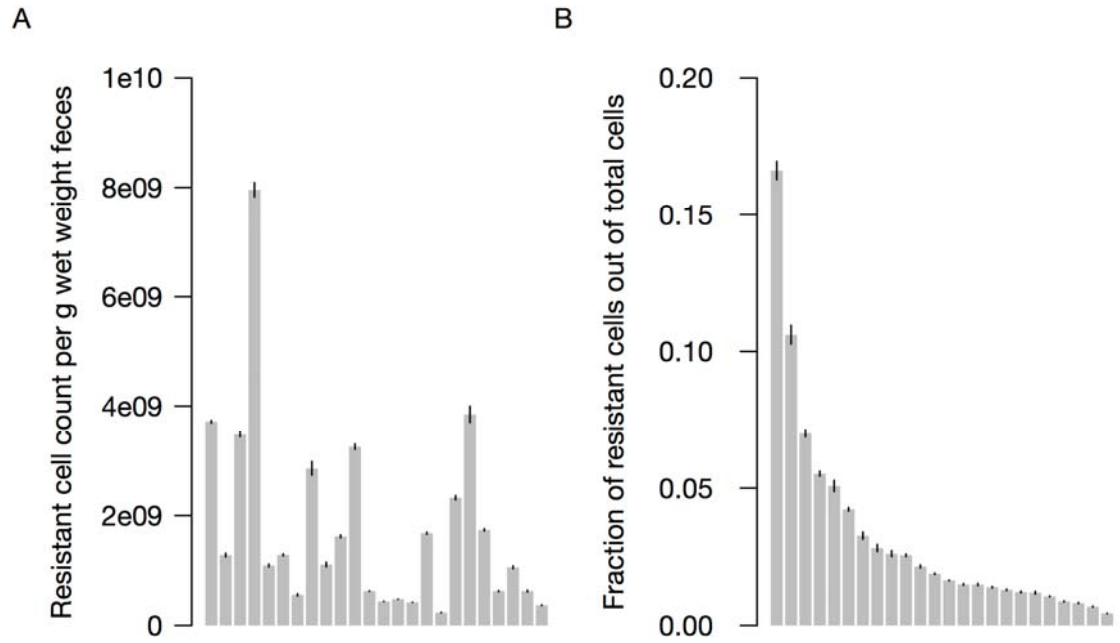
within the gastrointestinal tract permits sporulation for a number of these organisms as well, giving them an exit strategy if host conditions are sub-optimal.

We demonstrate that even though resistant cells themselves appear to be lowly abundant in human fecal material (on the order of 0.1-10% total biomass), sharing of rOTUs and the Clostridia in particular is favored across unrelated individuals. The increased sharing comes at the cost of vegetative cell abundance and persistence within a host. Because of heterogeneity in the resistance phenotype even within strains of the same species, few OTUs consistently form resistant cells at high levels across all individuals in our dataset. This heterogeneity drives a signal where organisms forming these resistant cells within any given person may not be doing so in other individuals, yet when considering endospore-forming potential as a whole (i.e. through its phylogenetic conservation within the Clostridia), OTUs with organisms that potentially can form endospores are more likely than others to be found among many unrelated individuals. Within a person, the organisms we observe forming resistant cells exhibit distinct correlated behavior coupled to the dynamic behavior of select non-endospore-forming taxa that likely respond to shared environmental signals.

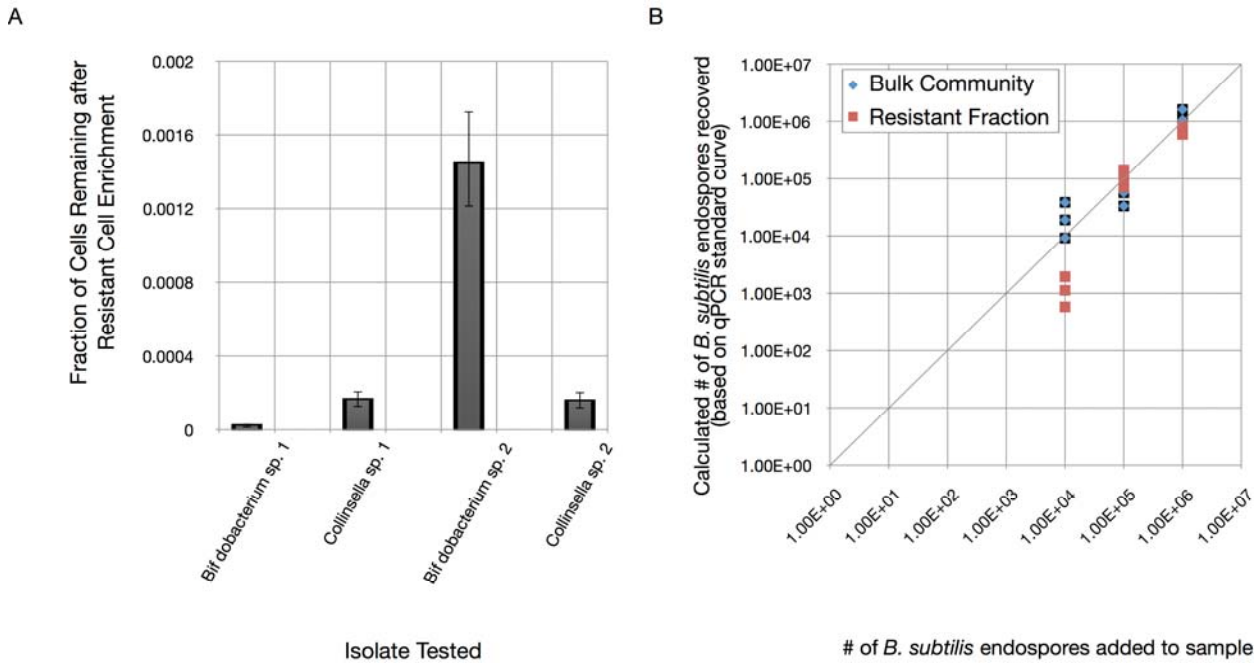


Supplementary Figure 1. Resistant OTUs are more prevalent across individuals in their bulk communities. The prevalence of OTUs across 24 individuals in our cohort as a function of fraction of times the OTU was enriched in the resistant fraction relative to the bulk community. The area of a point corresponds to the number of observations of OTUs at a given fractional enrichment, fractional sharing pair. Pink points correspond to nOTUs (never enriched in the endospore fraction) and blue points correspond to rOTUs (those ever enriched in the endospore fraction). The histogram on top of the plot relates the density of observations at each fractional enrichment level; the histogram on the side of the plot relates the density of observations for nOTUs (pink) and rOTUs (blue) across the sharing gradient.

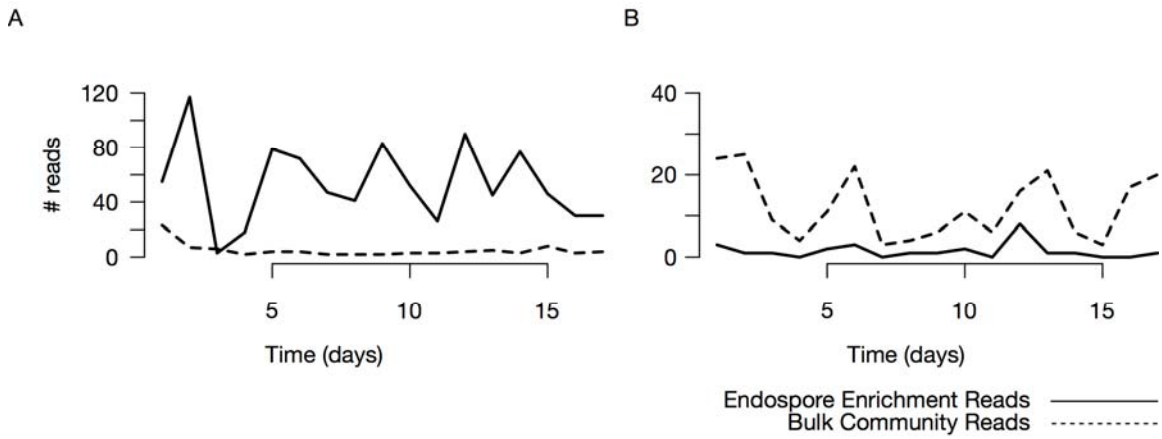
1062



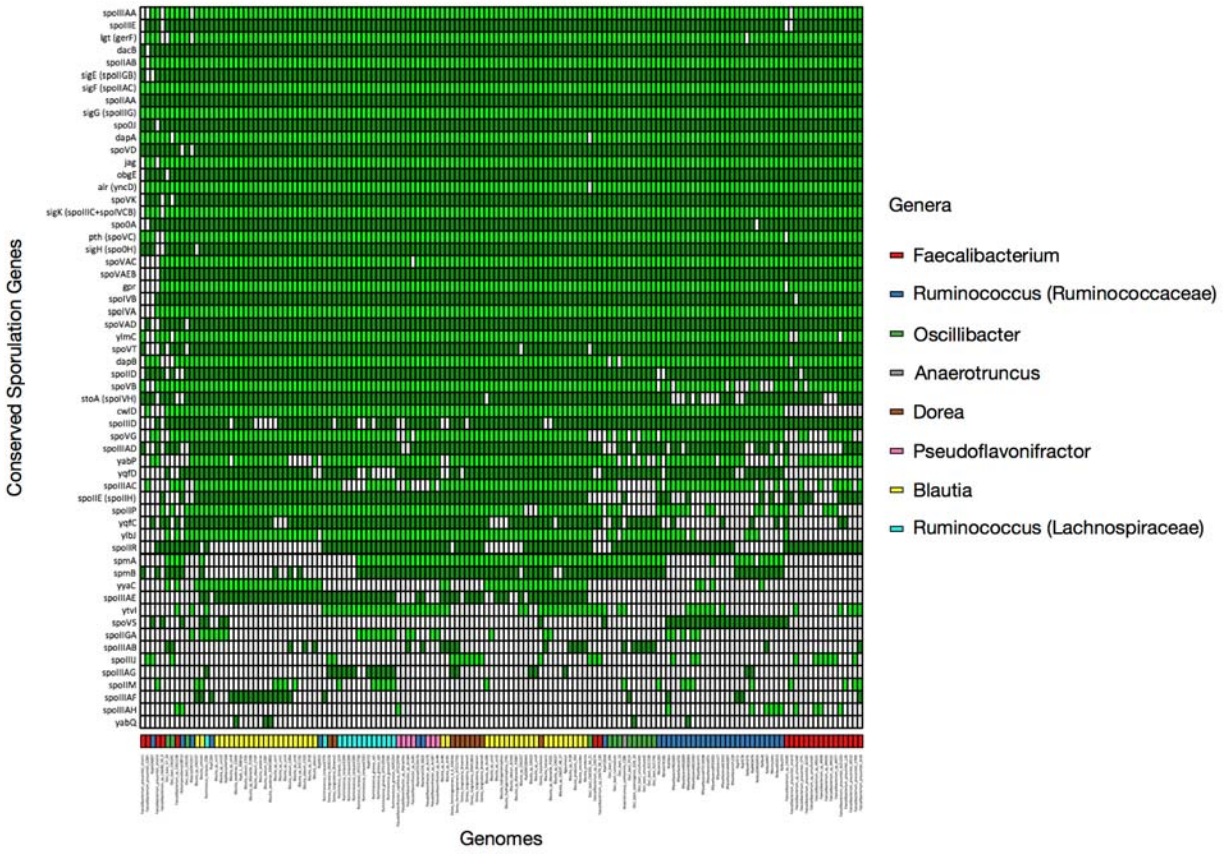
Supplementary Figure 2. Resistant cells generally comprise fewer than 2% of cells in the gut microbiota (A) Resistant cell abundance sorted by fraction of total cells estimated using qPCR on the resistant fraction quantified with a DNA standard curve of a mock community consisting of cultured isolates of *E. coli*, *L. reuteri*, *V. parvula*, *B. fragilis*, and *C. difficile*. (B) Endospore relative abundance as the fraction of total cells estimated using qPCR, sorted from highest to lowest.



Supplementary Figure 3. Cultured Actinobacteria isolated from the human gut are significantly depleted when enriching for resistant cells, but endospores are reliably recovered. (A) Fraction of DNA remaining after resistant cell enrichment is applied to cultured Actinobacteria. Fractions are estimated by qPCR for four human gut-derived isolates from the phylum Actinobacteria calculated by comparing the threshold cycle (Ct) for amplification of DNA using the V4 16S rDNA primers used for library assembly. Ct is converted to DNA concentration using a standard curve of DNA dilutions for each isolate. Error bars are the standard error of the mean. (B) Estimate of the recovery of endospores of *B. subtilis* added at different amounts into independent fecal samples (individual points) and treated with standard DNA extraction (blue diamonds) or resistant fraction (red squares). Endospore recovery is estimated using qPCR with primers targeting *aprE* (*aprE* F: 5'-TTTACGATGGCGTTCAGCAAC-3', *aprE* R: 5'-GGAAGTGCCTTCATTTCCGGCT-3')[63]. The line y = x indicates stoichiometric recovery, and the horizontal dashed line shows the limit of detection of the primer set.



Supplementary Figure 4. Time series of two distinct *Ruminococcus* OTUs show discordance in resistant fraction tendency. (A) Bulk community and resistant fraction time series for a *Ruminococcus* OTU with corresponding V4 16S region matching at 100% sequence identity to *Ruminococcus lactaris* ATCC 29176. (B) Bulk community and resistant fraction time series for a *Ruminococcus* OTU with corresponding V4 16S region matching at 97% sequence identity to *Ruminococcus faecis* strain Eg2.



Supplementary Figure 5. Variably endospore-enriched taxa lack several genes shared between previously identified endospore-forming Bacilli and Clostridia. Sporulation genes (y-axis) considered conserved between all endospore-forming *Bacilli* and *Clostridia*[31] were mapped to genomes of genera (using tblastn with an E-value cutoff of 1e-5) containing OTUs observed as significantly increased and decreased in the resistant fraction across individuals measured by Mann Whitney U test (FDR adjusted p-value < 0.10). Genomes on the x-axis are all genomes corresponding to the named genera downloaded from the NCBI database as of August 2017. Color bar on the x-axis corresponds to genera. Genomes are clustered using UPGMA hierarchical clustering on presence-absence profiles of sporulation gene content; genes profiles

1105 across genomes are clustered using the same procedure. Anaerotruncus colihominins

1106 (gray) is a known endospore-former that lacks several putatively conserved genes.

Supplementary Table 3. Significant (unadjusted p < 0.05) Spearman Correlations between metabolites and rOTU cluster.

Metabolite Name	m/z	Spearman Rho	Unadjusted p-value	FDR q-value
sebacate	201.1119	0.74	5.64E-05	3.89E-03
arachidonate	303.2321	-0.593	2.72E-03	3.13E-02
docosapentaenoate	329.2476	-0.582	3.39E-03	3.13E-02
eicosapentaenoate	301.2164	-0.58	3.51E-03	3.13E-02
nonadecanoate	297.279	-0.578	3.63E-03	3.13E-02
pentadecanoate	241.2162	0.591	2.81E-03	3.13E-02
3-oxooctadecanoate	297.2427	0.619	1.58E-03	3.13E-02
docosahexaenoate	327.232	-0.565	4.63E-03	3.55E-02
hydrocinnamate	149.0592	0.558	5.25E-03	3.62E-02
undecanedionate	215.1276	0.617	1.67E-03	3.62E-02
taurocholate	514.2825	-0.549	6.23E-03	3.91E-02
crustecdysone	525.3053	-0.608	2.37E-02	1.09E-01
adrenate	331.2633	-0.474	2.04E-02	1.09E-01
10-nonadecanoate	295.2633	-0.473	2.07E-02	1.09E-01
dimethylurate	195.0511	0.469	2.23E-02	1.09E-01
eicosatrienoate	305.2477	-0.438	3.33E-02	1.44E-01
arachidate	311.2947	-0.424	3.99E-02	1.62E-01
urobilin	593.3324	0.417	4.35E-02	1.67E-01

REFERENCES

1. Filippidou, S. *et al.* Under-detection of endospore-forming Firmicutes in metagenomic data. *Comput. Struct. Biotechnol. J.* **13,** 299–306 (2015).
2. Browne, H. P. *et al.* Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. *Nature* **533,** 543–546 (2016).
3. Ley, R. E. *et al.* Evolution of Mammals and Their Gut Microbes. *Science* **320,** 1647–1651 (2008).
4. Angert, E. R. & Losick, R. M. Propagation by sporulation in the guinea pig symbiont Metabacterium polyspora. *Proc. Natl. Acad. Sci.* **95,** 10218–10223 (1998).
5. Flint, J. F., Drzymalski, D., Montgomery, W. L., Southam, G. & Angert, E. R. Nocturnal Production of Endospores in Natural Populations of Epulopiscium-Like Surgeonfish Symbionts. *J. Bacteriol.* **187,** 7460–7470 (2005).
6. Alexander, C. J., Citron, D. M., Brazier, J. S. & Goldstein, E. J. Identification and antimicrobial resistance patterns of clinical isolates of Clostridium clostridioforme, Clostridium innocuum, and Clostridium ramosum compared with those of clinical isolates of Clostridium perfringens. *J. Clin. Microbiol.* **33,** 3209–3215 (1995).
7. Paredes-Sabja, D., Torres, J. A., Setlow, P. & Sarker, M. R. Clostridium perfringens Spore Germination: Characterization of Germinants and Their Receptors. *J. Bacteriol.* **190,** 1190–1201 (2008).
8. Sokol, H. *et al.* Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc. Natl. Acad. Sci.* **105,** 16731–16736 (2008).

9. Png, C. W. *et al.* Mucolytic bacteria with increased prevalence in IBD mucosa augment in vitro utilization of mucin by other bacteria. *Am. J. Gastroenterol.* **105,** 2420–2428 (2010).
10. Deakin, L. J. *et al.* The *Clostridium difficile spo0A* Gene Is a Persistence and Transmission Factor. *Infect. Immun.* **80,** 2704–2711 (2012).
11. Atarashi, K. *et al.* Induction of Colonic Regulatory T Cells by Indigenous Clostridium Species. *Science* **331,** 337–341 (2011).
12. Atarashi, K. *et al.* Treg induction by a rationally selected mixture of Clostridia strains from the human microbiota. *Nature* **500,** 232–236 (2013).
13. Stefka, A. T. *et al.* Commensal bacteria protect against food allergen sensitization. *Proc. Natl. Acad. Sci.* **111,** 13145–13150 (2014).
14. Kim, Y.-G. *et al.* Neonatal acquisition of Clostridia species protects against colonization by bacterial pathogens. *Science* **356,** 315 (2017).
15. Ivanov, I. I. *et al.* Induction of Intestinal Th17 Cells by Segmented Filamentous Bacteria. *Cell* **139,** 485–498 (2009).
16. Kuwahara, T. *et al.* The Lifestyle of the Segmented Filamentous Bacterium: A Non-Culturable Gut-Associated Immunostimulating Microbe Inferred by Whole-Genome Sequencing. *DNA Res.* **18,** 291–303 (2011).
17. Sczesnak, A. *et al.* The Genome of Th17 Cell-Inducing Segmented Filamentous Bacteria Reveals Extensive Auxotrophy and Adaptations to the Intestinal Environment. *Cell Host Microbe* **10,** 260–272 (2011).
18. Smith, P. M. *et al.* The Microbial Metabolites, Short-Chain Fatty Acids, Regulate Colonic Treg Cell Homeostasis. *Science* **341,** 569–573 (2013).

19. Furusawa, Y. *et al.* Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature* **504,** 446–450 (2013).

20. Van den Abbeele, P. *et al.* Butyrate-producing Clostridium cluster XIVa species specifically colonize mucins in an in vitro gut model. *ISME J.* **7,** 949–961 (2013).

21. Eeckhaut, V. *et al.* Butyrate production in phylogenetically diverse Firmicutes isolated from the chicken caecum: Butyrate-producing bacteria from the chicken caecum. *Microb. Biotechnol.* **4,** 503–512 (2011).

22. Louis, P., Young, P., Holtrop, G. & Flint, H. J. Diversity of human colonic butyrate-producing bacteria revealed by analysis of the butyryl-CoA:acetate CoA-transferase gene. *Environ. Microbiol.* **12,** 304–314 (2010).

23. Yano, J. M. *et al.* Indigenous Bacteria from the Gut Microbiota Regulate Host Serotonin Biosynthesis. *Cell* **161,** 264–276 (2015).

24. Kaplan, I. & Williams, J. W. Spore Formation among the Anaerobic Bacteria: I. The Formation of Spores by Clostridium sporogenes in Nutrient Agar Media. *J. Bacteriol.* **42,** 265 (1941).

25. Dingman, D. W. & Stahly, D. P. Medium promoting sporulation of Bacillus larvae and metabolism of medium components. *Appl. Environ. Microbiol.* **46,** 860–869 (1983).

26. Wunderlin, T., Junier, T., Roussel-Delif, L., Jeanneret, N. & Junier, P. Endospore-enriched sequencing approach reveals unprecedented diversity of *Firmicutes* in sediments: Endospore-forming enrichment. *Environ. Microbiol. Rep.* **6,** 631–639 (2014).

1177 27. Sekar, R. *et al.* An Improved Protocol for Quantification of Freshwater Actinobacteria
1178 by Fluorescence In Situ Hybridization. *Appl. Environ. Microbiol.* **69,** 2928–2935
1179 (2003).
1180 28. Fahlgren, A., Hammarström, S., Danielsson, Å. & HAMMARSTRÖM, M.-L.
1181 Increased expression of antimicrobial peptides and lysozyme in colonic epithelial
1182 cells of patients with ulcerative colitis. *Clin. Exp. Immunol.* **131,** 90–101 (2003).
1183 29. Keshav, S., Chung, P., Milon, G. & Gordon, S. Lysozyme is an inducible marker of
1184 macrophage activation in murine tissues as demonstrated by in situ hybridization. *J.*
1185 *Exp. Med.* **174,** 1049 (1991).
1186 30. Gueimonde, M., Laitinen, K., Salminen, S. & Isolauri, E. Breast milk: a source of
1187 bifidobacteria for infant gut development and maturation? *Neonatology* **92,** 64–66
1188 (2007).
1189 31. Galperin, M. Y. *et al.* Genomic determinants of sporulation in *Bacilli* and *Clostridia*□:
1190 towards the minimal set of sporulation-specific genes: Distribution of sporulation
1191 genes in *Bacilli* and *Clostridia*. *Environ. Microbiol.* **14,** 2870–2890 (2012).
1192 32. Popham, D. L., Illades-Aguiar, B. & Setlow, P. The Bacillus subtilis dacB gene,
1193 encoding penicillin-binding protein 5*, is part of a three-gene operon required for
1194 proper spore cortex synthesis and spore core dehydration. *J. Bacteriol.* **177,** 4721–
1195 4729 (1995).
1196 33. Orsburn, B., Sucre, K., Popham, D. L. & Melville, S. B. The SpmA/B and DacF
1197 proteins of Clostridium perfringens play important roles in spore heat resistance.
1198 *FEMS Microbiol. Lett.* **291,** 188–194 (2009).

34. Serra, C. R., Earl, A. M., Barbosa, T. M., Kolter, R. & Henriques, A. O. Sporulation during Growth in a Gut Isolate of Bacillus subtilis. *J. Bacteriol.* **196,** 4184–4196 (2014).

35. Francis, M. B., Allen, C. A., Shrestha, R. & Sorg, J. A. Bile Acid Recognition by the Clostridium difficile Germinant Receptor, CspC, Is Important for Establishing Infection. *PLoS Pathog.* **9,** e1003356 (2013).

36. Devkota, S. *et al.* Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in Il10−/− mice. *Nature* (2012). doi:10.1038/nature11225

37. Sorg, J. A. & Sonenshein, A. L. Inhibiting the Initiation of Clostridium difficile Spore Germination using Analogs of Chenodeoxycholic Acid, a Bile Acid. *J. Bacteriol.* **192,** 4983–4990 (2010).

38. Ceuppens, S. *et al.* Survival and Germination of Bacillus cereus Spores without Outgrowth or Enterotoxin Production during In Vitro Simulation of Gastrointestinal Transit. *Appl. Environ. Microbiol.* **78,** 7698–7705 (2012).

39. Shah, I. M., Laaberki, M.-H., Popham, D. L. & Dworkin, J. A eukaryotic-like Ser/Thr kinase signals bacteria to exit dormancy in response to peptidoglycan fragments. *Cell* **135,** 486–496 (2008).

40. Koenig, J. E. *et al.* Succession of microbial consortia in the developing infant gut microbiome. *Proc. Natl. Acad. Sci.* **108,** 4578–4585 (2011).

41. Chandan, R. C., Shahani, K. M. & Holly, R. G. Lysozyme Content of Human Milk. *Nature* **204,** 76 (1964).

1220 42. Nayfach, S., Rodriguez-Mueller, B., Garud, N. & Pollard, K. S. An integrated
1221 metagenomics pipeline for strain profiling reveals novel patterns of bacterial
1222 transmission and biogeography. *Genome Res.* **26,** 1612–1625 (2016).
1223 43. Allegretti, J. R. *et al.* Recurrent *Clostridium difficile* infection associates with distinct
1224 bile acid and microbiome profiles. *Aliment. Pharmacol. Ther.* **43,** 1142–1153 (2016).
1225 44. Youngster, I. *et al.* Fecal Microbiota Transplant for Relapsing Clostridium difficile
1226 Infection Using a Frozen Inoculum From Unrelated Donors: A Randomized, Open-
1227 Label, Controlled Pilot Study. *Clin. Infect. Dis.* **58,** 1515–1522 (2014).
1228 45. David, L. A. *et al.* Host lifestyle affects human microbiota on daily timescales.
1229 *Genome Biol* **15,** R89 (2014).
1230 46. Bueche, M. *et al.* Quantification of Endospore-Forming Firmicutes by Quantitative
1231 PCR with the Functional Gene spo0A. *Appl. Environ. Microbiol.* **79,** 5302–5312
1232 (2013).
1233 47. Chomczynski, P. & Rymaszewski, M. Alkaline polyethylene glycol-based method for
1234 direct PCR from bacteria, eukaryotic tissue samples, and whole blood.
1235 *Biotechniques* **40,** 454 (2006).
1236 48. Preheim, S. P., Perrotta, A. R., Martin-Platero, A. M., Gupta, A. & Alm, E. J.
1237 Distribution-Based Clustering: Using Ecology To Refine the Operational Taxonomic
1238 Unit. *Appl. Environ. Microbiol.* **79,** 6593–6603 (2013).
1239 49. Fierer, N., Jackson, J. A., Vilgalys, R. & Jackson, R. B. Assessment of soil microbial
1240 community structure by use of taxon-specific quantitative PCR assays. *Appl.*
1241 *Environ. Microbiol.* **71,** 4117–4120 (2005).

50. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30,** 614–620 (2014).

51. Maidak, B. L. *et al.* The ribosomal database project (RDP). *Nucleic Acids Res.* **24,** 82–85 (1996).

52. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73,** 5261–5267 (2007).

53. Paulson, J. N., Stine, O. C., Bravo, H. C. & Pop, M. Differential abundance analysis for microbial marker-gene surveys. *Nat. Methods* **10,** 1200–1202 (2013).

54. Nicholson, W. & Setlow, P. Purification of spores on step gradients of Renografin. in *Molecular Biological Methods for Bacillus* 391–450 (John Wiley, 1990).

55. Fichtel, J., Köster, J., Rullkötter, J. & Sass, H. Spore dipicolinic acid contents used for estimating the number of endospores in sediments: Determination of endospores in sediments. *FEMS Microbiol. Ecol.* **61,** 522–532 (2007).

56. Khanna, S. *et al.* A Novel Microbiome Therapeutic Increases Gut Microbial Diversity and Prevents Recurrent *Clostridium difficile* Infection. *J. Infect. Dis.* **214,** 173–181 (2016).

57. Lopez-Siles, M. *et al.* Cultured Representatives of Two Major Phylogroups of Human Colonic Faecalibacterium prausnitzii Can Utilize Pectin, Uronic Acids, and Host-Derived Substrates for Growth. *Appl. Environ. Microbiol.* **78,** 420–428 (2012).

58. Lopetuso, L. R., Scaldaferri, F., Petito, V. & Gasbarrini, A. Commensal Clostridia: leading players in the maintenance of gut homeostasis. *Gut Pathog.* **5,** 1 (2013).

59. Smedman, A. E., Gustafsson, I.-B., Berglund, L. G. & Vessby, B. O. Pentadecanoic acid in serum as a marker for intake of milk fat: relations between intake of milk fat and metabolic risk factors. *Am. J. Clin. Nutr.* **69,** 22–29 (1999).

60. Hayashi, A. *et al.* A Single Strain of Clostridium butyricum Induces Intestinal IL-10-Producing Macrophages to Suppress Acute Experimental Colitis in Mice. *Cell Host Microbe* **13,** 711–722 (2013).

61. Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Comprehensive comparative-genomic analysis of Type 2 toxin-antitoxin systems and related mobile stress response systems in prokaryotes. *Biol. Direct* **4,** 19 (2009).

62. Maurice, C. F., Haiser, H. J. & Turnbaugh, P. J. Xenobiotics Shape the Physiology and Gene Expression of the Active Human Gut Microbiome. *Cell* **152,** 39–50 (2013).

63. Sadeghi, A., Mortazavi, S. A., Bahrami, A. R. & Sadeghi, B. Design of multiplex PCR for simultaneous detection of rope‐forming Bacillus strains in Iranian bread dough. *J. Sci. Food Agric.* **92,** 2652–2656 (2012).