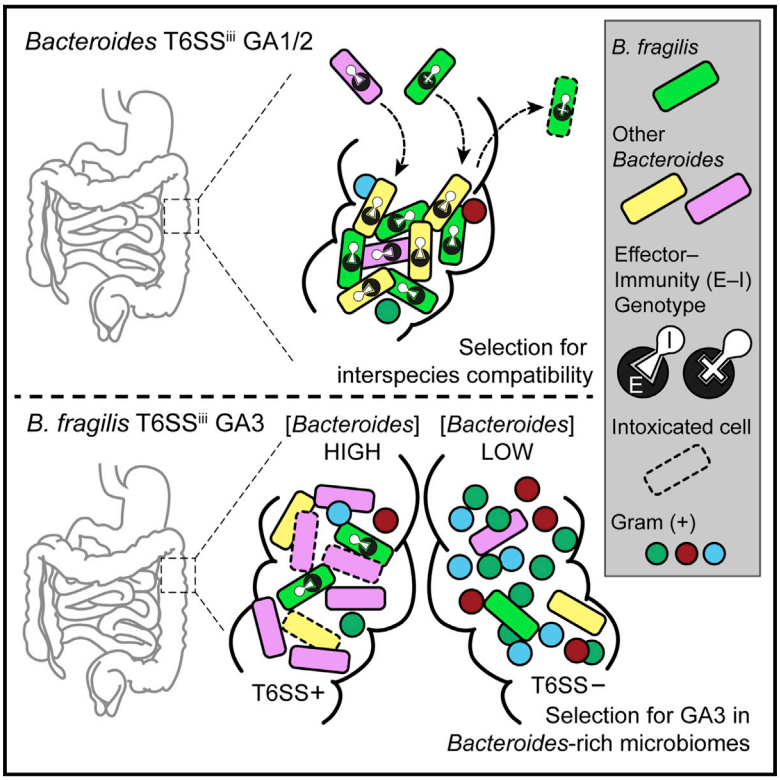
Short Article

Cell Host & Microbe

The Landscape of Type VI Secretion across Human Gut Microbiomes Reveals Its Role in Community Composition

Graphical Abstract



Highlights

- Comprehensive metagenomic mapping of *Bacteroidales* T6SS genes in the gut microbiome
- Evidence for selection for T6SS compatibility at the species and strain levels
- *Bacteroides fragilis* strains compete for dominance in the infant microbiome
- *B. fragilis* T6SS provides a selective advantage in *Bacteroides*-rich microbiomes

Authors

Adrian J. Verster, Benjamin D. Ross, Matthew C. Radey, Yiqiao Bao, Andrew L. Goodman, Joseph D. Mougous, Elhanan Borenstein

Correspondence

mougous@uw.edu (J.D.M.), elbo@uw.edu (E.B.)

In Brief

The T6SS is an effector delivery system that mediates interbacterial competition. Using metagenomic analyses, Verster et al. investigate the prevalence and role of the T6SS in the human gut microbiome. They demonstrate that the T6SS mediates interactions between *Bacteroides* strains in the infant microbiome and between species in *Bacteroides*-rich environments.

Verster et al., 2017, Cell Host & Microbe *22*, 411–419
September 13, 2017
http://dx.doi.org/10.1016/j.chom.2017.08.010

Cell Host & Microbe
Short Article

The Landscape of Type VI Secretion across Human Gut Microbiomes Reveals Its Role in Community Composition

Adrian J. Verster,[1,8] Benjamin D. Ross,[2,8] Matthew C. Radey,[2] Yiqiao Bao,[3,4] Andrew L. Goodman,[3,4] Joseph D. Mougous,[2,5,*] and Elhanan Borenstein[1,6,7,9,*]
[1]Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA
[2]Department of Microbiology, School of Medicine, University of Washington, Seattle, WA 98195, USA
[3]Department of Microbial Pathogenesis, Yale University School of Medicine, New Haven, CT 06510, USA
[4]Microbial Sciences Institute, Yale University School of Medicine, West Haven, CT 06516, USA
[5]Howard Hughes Medical Institute, School of Medicine, University of Washington, Seattle, WA 98195, USA
[6]Department of Computer Science and Engineering, University of Washington, Seattle, WA 98195, USA
[7]Santa Fe Institute, Santa Fe, NM 87501, USA
[8]These authors contributed equally
[9]Lead Contact
*Correspondence: mougous@uw.edu (J.D.M.), elbo@uw.edu (E.B.)
http://dx.doi.org/10.1016/j.chom.2017.08.010

SUMMARY

Although gut microbiome composition is well defined, the mechanisms underlying community assembly remain poorly understood. *Bacteroidales* possess three genetic architectures (GA1–3) of the type VI secretion system (T6SS), an effector delivery pathway that mediates interbacterial competition. Here we define the distribution and role of GA1–3 in the human gut using metagenomic analysis. We find that adult microbiomes harbor limited effector and cognate immunity genes, suggesting selection for compatibility at the species (GA1 and GA2) and strain (GA3) levels. *Bacteroides fragilis* GA3 is known to mediate potent inter-strain competition, and we observe GA3 enrichment among strains colonizing infant microbiomes, suggesting competition early in life. Additionally, GA3 is associated with increased *Bacteroides* abundance, indicating that this system confers an advantage in *Bacteroides*-rich ecosystems. Collectively, these analyses uncover the prevalence of T6SS-dependent competition and reveal its potential role in shaping human gut microbial composition.

INTRODUCTION

Bacterial communities are of fundamental importance to natural ecosystems (Prosser et al., 2007). While cooperative interactions between the species comprising such communities can occur (Rakoff-Nahoum et al., 2016), it is clear that bacteria in these settings experience pervasive competition from surrounding cells (Coyte et al., 2015; Hibbing et al., 2010; Levy and Borenstein, 2013). Indeed, the genomes of bacteria encode a wealth of dedicated interbacterial antagonism pathways (Zhang et al., 2012). Some of these function through the production of diffusible small molecules (Riley and Wertz, 2002), whereas others utilize proteinaceous toxins. A prevalent pathway mediating the transfer of toxic proteins between bacteria is the type VI secretion system (T6SS) (Hood et al., 2010). This system has been most thoroughly studied in *Proteobacteria*, though it is found in several phyla of Gram-negative bacteria, and while it can, in some cases, target eukaryotic cells, it has been primarily investigated as an interbacterial system.

The T6S apparatus transfers toxic effector proteins from donor to recipient bacterial cells by a mechanism dependent upon cell contact (Russell et al., 2014a). Characterized interbacterial effector proteins are thus far without exception enzymes that target conserved, essential features of the bacterial cell, such as peptidoglycan, phospholipids, and nucleic acids. This feature of effector proteins, taken together with the fact that T6SS targeting does not appear to be dependent on a specific receptor, confers broad activity against Gram-negative cells. Indiscriminate effector transfer also extends to kin cells; therefore, cells with the T6SS produce immunity proteins that inactivate cognate toxins through active site occlusion (Benz and Meinhart, 2014).

Given its wide phylogenetic distribution and its capacity to target diverse recipient cells, the T6SS is likely to play an important role in the assembly and composition of bacterial communities. Indeed, there are recent reports consistent with the pathway mediating bacterial interaction in environmental communities. For instance, T6S genes were found to be enriched and under positive selection in the barley rhizosphere (Bulgarelli et al., 2015), and T6S phospholipase effectors were detected in metagenomes from diverse sources (Egan et al., 2015). To date, however, systematic studies of the impact of T6S on microbial community assembly are lacking.

The human gut microbiome is a dense ecosystem whose composition is paramount to its function (Walter and Ley, 2011). Factors such as diet, immune status, and host genetics

have each been implicated in shaping the gut community, yet the contribution of direct interbacterial competition to the structure of this community remains poorly understood. Recently, a T6SS-like pathway was detected in *Bacteriodetes*, the most abundant Gram-negative phylum in the human gut (Coyne et al., 2014; Russell et al., 2014b). Additional work demonstrated that T6S contributes to the fitness of *Bacteroides fragilis* in competition with other bacteria in vitro and in gnotobiotic mice (Chatzidaki-Livanis et al., 2016; Hecht et al., 2016; Russell et al., 2014b; Wexler et al., 2016). These and other data show that the mammalian GI tract is physically conducive to T6SS-dependent interbacterial antagonism, suggesting a potential impact of this pathway on the composition of the human gut microbiome (Anderson et al., 2017; Sana et al., 2016). Here, we sought to define the distribution of the *Bacteroidales* T6SS and to explore its function in the human gut microbiome through the analysis of several publicly available metagenomic datasets. These datasets allow us to study the outcome of natural community dynamics in the gut microbiome, and we reasoned that their analysis could therefore provide unique insight into the physiologic role of T6SS-dependent competition in this ecosystem. Our findings reveal the prevalence of this pathway in intact human gut microbial communities, highlight striking and non-random patterns in its distribution across samples, and suggest an active role for the T6SS in intra- and inter-species bacterial interactions in the gut.

RESULTS

Detection of T6SS E–I Pairs in the Human Gut Microbiome

We first set out to characterize the prevalence and distribution of T6SS genes in the gut microbiomes of healthy adult individuals. Based on their organization and content, *Bacteroidales* T6SS gene clusters can be divided into three distinct subtypes, termed genetic architecture 1–3 (GA1–3) (Coyne et al., 2016). Each T6S subtype possesses one or more cassettes at stereotyped positions that contain variable genes predicted or demonstrated to encode effector–immunity (E–I) pairs (Chatzidaki-Livanis et al., 2016; Coyne et al., 2016; Russell et al., 2014b; Wexler et al., 2016). As T6SS-based antagonism is determined by the effector and immunity genes of donor and recipient cells, respectively, the identification of E–I pairs provides information regarding the potential for interbacterial interactions mediated by this system. Furthermore, since these cassettes are variable within, but appear unique among, the T6S subtypes, estimation of the abundance of these genes within metagenomes can serve as a proxy for the presence and distribution of GA1–3.

To define the E–I repertoire associated with GA1–3, we searched within T6-associated variable cassettes from *Bacteroidales* reference genomes and from pre-assembled metagenome contigs from the Human Microbiome Project (HMP) for genes with hallmarks of known T6SS effector and immunity factors (Human Microbiome Project Consortium, 2012). These included fusion to modular adaptor domains, reduced GC content, bicistronic arrangement, and similarity to protein families defined by their association with characterized E–I pairs (see STAR Methods for a complete description of annotation criteria; Figures 1A and S1). In total, we identified 12 GA1, 19 GA2, and 14 GA3 putative E–I pairs. As expected, genes with significant homology to GA3 pairs were identified only in *B. fragilis* reference genomes, whereas GA1 and GA2 pairs were detected throughout the order. Importantly, we did not identify GA1–3 E–I genes outside of *Bacteroidales*.

To estimate the abundance of these E–I pairs in gut microbiomes, we obtained metagenomic datasets derived from healthy donor samples of the HMP (Human Microbiome Project Consortium, 2012) and MetaHIT (Qin et al., 2010) studies. We next mapped the reads from each sample to our catalog of E–I genes (using 97% sequence identity threshold). Our results indicate that T6S is prevalent in the human gut microbiome; of the 246 samples analyzed, we detected E–I genes in 155 (63%). Moreover, each E–I pair in our list was detected in at least one microbiome sample, with an average of 9.5 occurrences. Importantly, the abundance of GA1–3 effector genes across samples correlates with that of subtype-specific T6SS structural genes with only very few samples containing structural genes and no effector genes, suggesting that our catalog of E–I pairs is comprehensive and approximates the full diversity of such genes in nature (Figure 2A).

T6S E–I Pairs Display Low Diversity within Human Gut Microbiome Samples

The systematic characterization of E–I abundance in metagenomic samples provided a unique opportunity to examine the distribution of the genes associated with each genetic architecture across healthy gut microbiomes. We first focused on GA1 and GA2, which utilize unique complements of effectors, but share the ability to undergo conjugative transfer between species belonging to the order *Bacteroidales* (Coyne et al., 2016). Surprisingly, we found that the complement of GA1- and GA2-associated E–I genes in a typical microbiome is small, with only a few pairs per sample, comparable to the number of pairs usually detected in a single genome (Figures 1B, 1C, and 1E). Moreover, in many cases, the same complement of E–I genes was detected in multiple samples. Henceforth, we refer to these combinations as E–I genotypes. This pattern suggests that either each sample is dominated by a single strain that harbors the observed E–I genotype or that there exists selective pressure for compatible E–I genes across multiple strains or species in a sample.

To further explore these possibilities, we focused our attention on the most prominent members of the genus *Bacteroides*. Other genera in the order *Bacteroidales* are less abundant constituents of the microbiome and based on reference genomes do not often harbor GA1 or GA2. We identified a set of species-specific single-copy marker genes for each *Bacteroides* species and estimated their abundance in each sample. Next we compared marker gene abundance to that of GA1 and GA2 E–I genes across samples (see STAR Methods). We found that the abundance of these E–I genes was not consistent with that of an individual species (Figures 2B and 2C). These findings suggest that multiple species co-existing in a microbiome typically encode a single GA1 and/or GA2 E–I genotype, potentially due to selective pressure for maintenance of E–I compatibility.

We next examined GA3 E–I genes and found, as in GA1 and GA2, that each sample harbors only a small set of E–I pairs (Figures 1D and 1E). Moreover, observed GA3 E–I genotypes

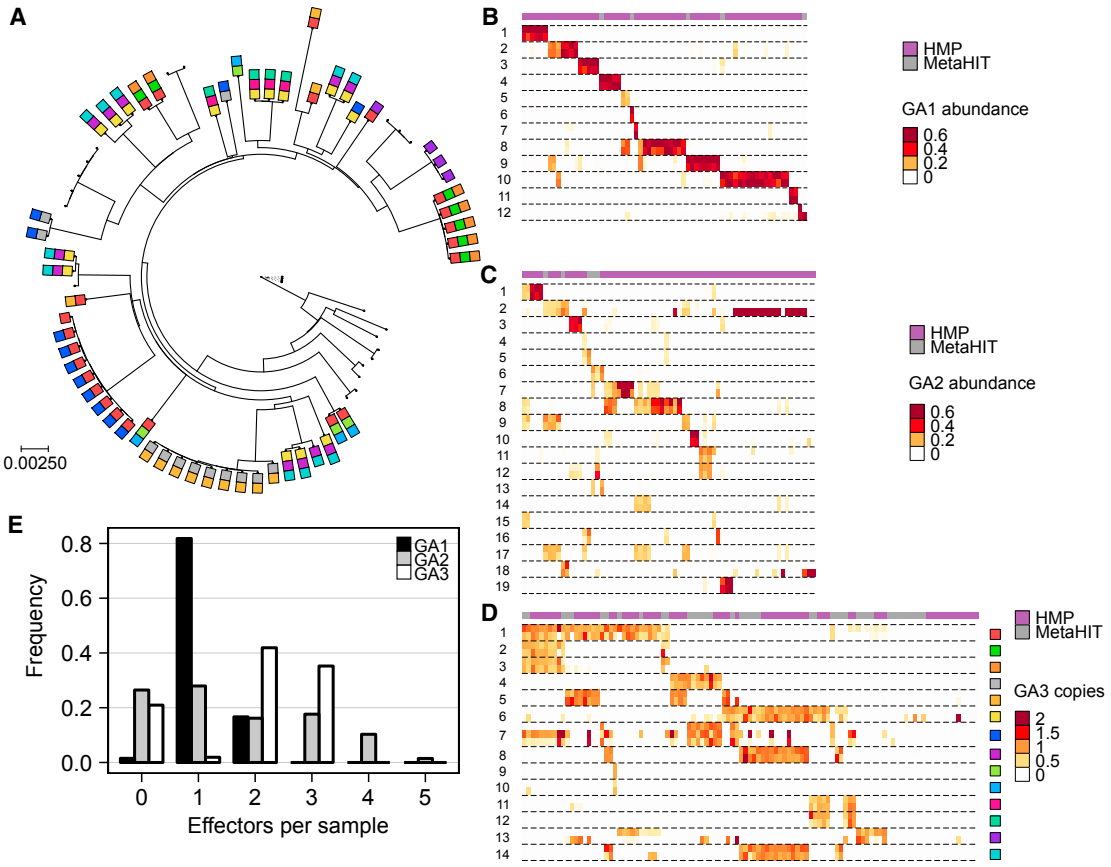


Figure 1. *Bacteroides* T6SS E–I Genes Are Abundant in Human Gut Microbiome Samples

(A) A maximum likelihood phylogeny of *B. fragilis* reference strains constructed from concatenated marker genes. Phylogenetic distance is measured as substitutions per site on the marker genes. GA3 effector genes are represented as colored squares (using the same color coding as in D).

(B–D) Each heatmap illustrates the abundance of E–I genes for one of the T6SS subsystems. Each row corresponds to a different E–I pair (effector, top; immunity, bottom). Columns represent the samples analyzed (HMP, purple; MetaHIT, gray). For GA1 (B) and GA2 (C), only samples in which at least 100 reads mapped to the E–I genes of a given subsystem are included, and abundance is measured as the fraction of the total abundance of E–I genes in a given sample. For GA3 (D), only samples in which *B. fragilis* is present are included and E–I abundance is normalized by the abundance of *B. fragilis*-specific marker genes, hence measuring the average number of copies per *B. fragilis* genome.

(E) Histograms showing the number of effector genes detected (at >10% of the most abundant effector gene) in each sample.

matched those detected in reference genomes (Figures S2A and S2B) and appeared randomly distributed between the American (HMP) and European (MetaHIT) datasets (Figure 1D). However, in contrast to GA1 and GA2, we found a strong correlation (R = 0.94) between the abundance of GA3 effector and immunity genes and that of a single species, *B. fragilis* (Figures 2B–2D). This finding shows that restriction of GA3 to *B. fragilis* observed in sequenced reference genomes holds across naturally occurring communities (Coyne et al., 2016).

We hypothesized that the pattern of GA3 E–I genotypes we observed could be explained by the dominance of a single *B. fragilis* strain within each individual microbiome. Indeed, prior studies suggest that *B. fragilis* exhibits relatively low diversity within individuals (Yassour et al., 2016). To confirm that this pattern is also observed in HMP and MetaHIT samples, we first measured nucleotide diversity in species-specific markers of *Bacteroides* spp. We found that within an individual, *B. fragilis* possesses the lowest average SNP diversity of well-represented members of the *Bacteroides* genus (Figure S2C). We then used a previously developed method for inferring the most likely set of strains in metagenomic samples based on nucleotide variants, combined with a phylogenetic analysis of these inferred strains, to determine the number of different monophyletic groups of strains present in each sample (STAR Methods). We found that *B. fragilis*-inferred strains in every HMP and MetaHIT sample formed a single monophyletic group, indicating that extant *B. fragilis* strains in each sample are likely derived from a single colonization event or outcompeted other strains to obtain dominance. Moreover, the set of E–I pairs detected in each metagenomic sample generally matched the set of E–I pairs found in the reference strains closest in the phylogenetic tree to the inferred strains, especially when the distance of inferred strains to their nearest reference was low (Figure S3).

To experimentally confirm our computational findings, we additionally selected 20 *B. fragilis* colonies isolated from two healthy adults and subjected these to whole-genome sequencing. Consistent with our findings using metagenomic data, our

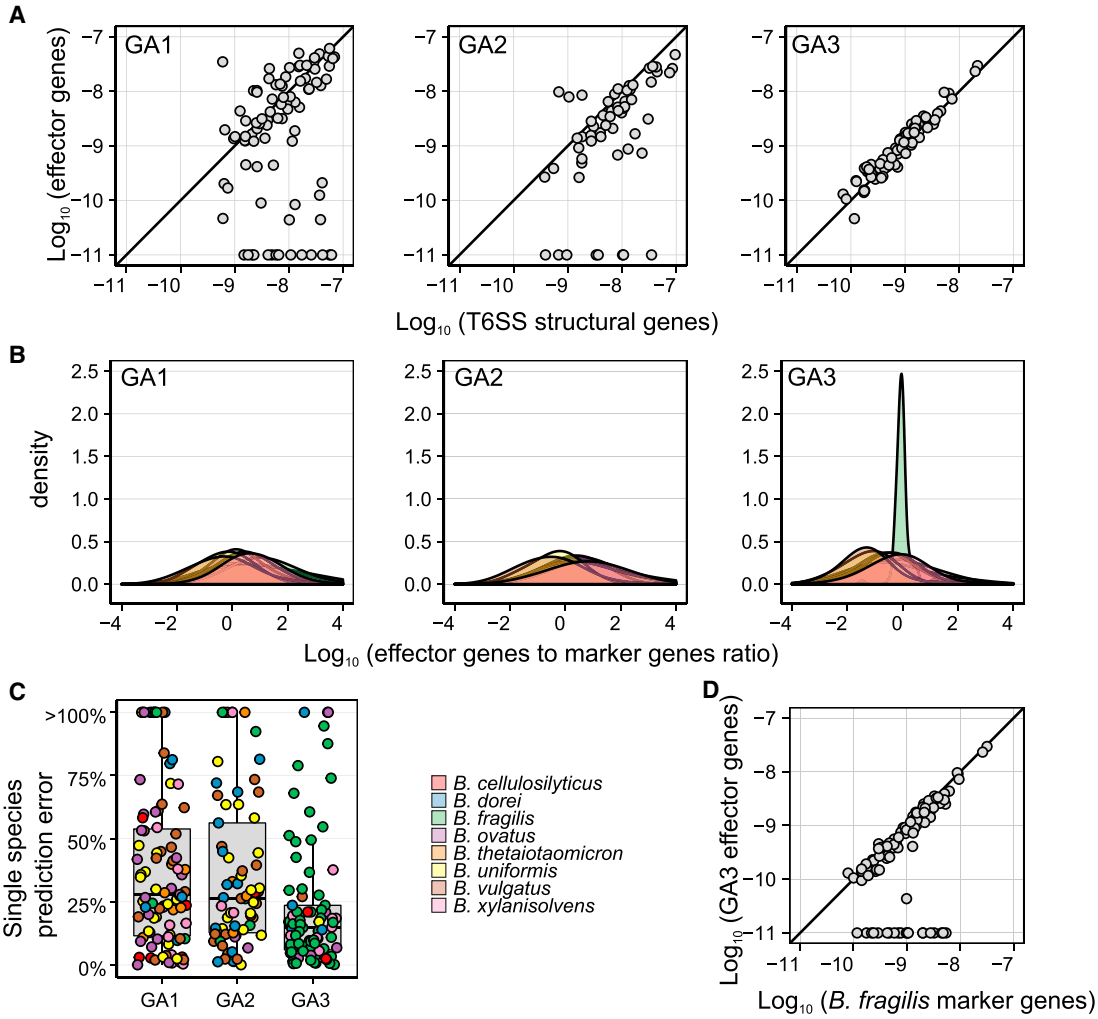


Figure 2. Differential Associations between T6SS and *Bacteroides* spp.

(A) Scatterplot of the average abundance of detected effector genes versus the average abundance of T6SS structural genes for different subtypes. We have restricted our analysis to samples with at least 25 reads mapping to a given subtype. The strong correlation observed, and the very few samples in which structural genes but no effector genes can be found, testify to the completeness of our E–I pairs catalog.

(B) Density plots showing the distribution across samples of the ratio between the average abundance of detected effector genes from each T6SS subsystem and the average abundance of species-specific marker genes for different *Bacteroides* spp. Only samples in which at least 100 reads mapped to the E–I genes of a given subsystem and only species for which at least five genomes were available (and therefore marker genes can be robustly inferred) are included.

(C) A boxplot showing the minimal relative error in effector abundance assuming that the T6SS is encoded by a single species. The relative error is defined as the relative difference between the average abundance of detectable effector genes in a sample and the abundance of species with the closest abundance. The color of each point represents the species for which the minimal relative error was obtained.

(D) Scatterplot of the average abundance of detected GA3 effector genes versus the average abundance of *B. fragilis*-specific marker genes. Only samples in which *B. fragilis* is present are included. As in (A), each abundance was increased by 10^{-11}.

See also Figures S2 and S3.

sequencing showed that a single clonal strain of *B. fragilis* dominates the microbiome of these individuals (Figure S2D).

B. fragilis GA3 Is Important in the Developing Microbiome

The finding that the presence of singular GA3 genotypes within individuals is due to the dominance of one *B. fragilis* strain motivated us to investigate the role of this system in the microbiome. We reasoned that in this dense and competitive microbiome ecosystem, an antagonistic pathway such as the T6SS might provide a fitness advantage (Ley et al., 2006). The system could mediate antagonism against other *B. fragilis* strains, *Bacteroides* spp., Gram-negative inhabitants of the microbiome, or a combination of these. Assuming such a role for T6SS, we further reasoned that in the microbiome of infants, which is less stable than that of adults, the function of an antagonistic pathway like the T6SS might be more pronounced. To test this hypothesis, we obtained publically available metagenomic datasets derived from infant gut microbiomes (Bäckhed et al., 2015; Kostic et al., 2015; Vatanen et al., 2016; Yassour et al., 2016). We then

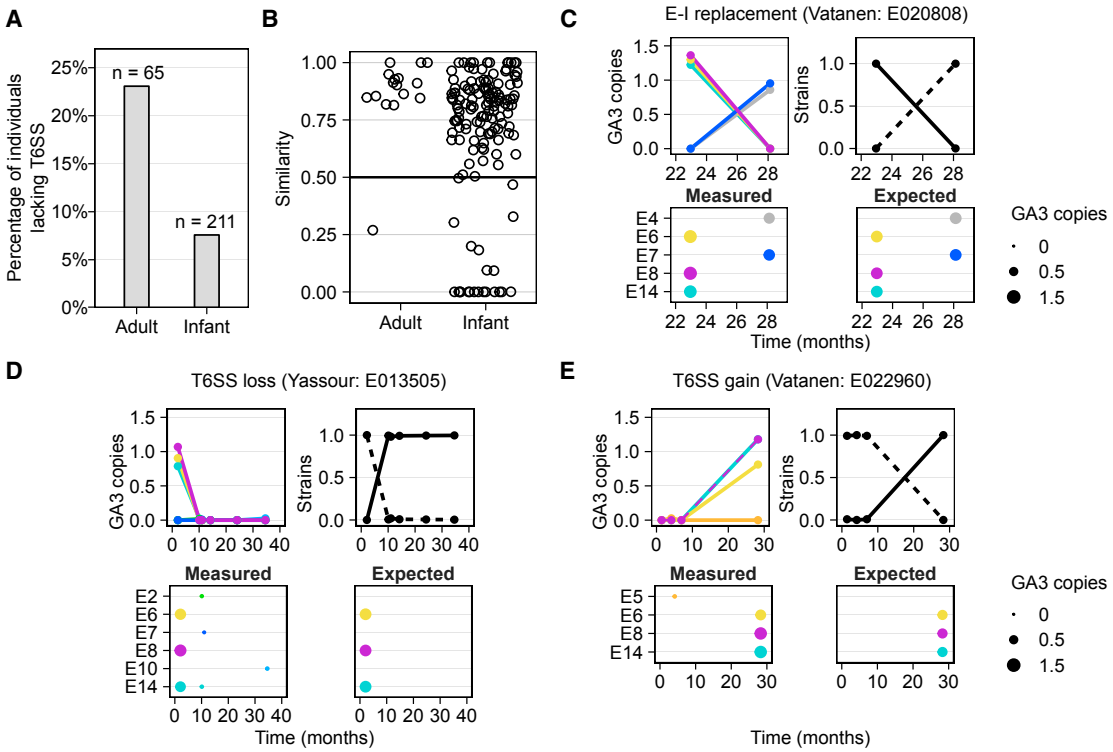


Figure 3. E–I Turnover and Strain Replacement in Infant Microbiomes

(A) The percentage of individuals of those harboring *B. fragilis* that lack the GA3 T6SS across adult and infant datasets.

(B) The minimal similarity (measured by the Jaccard similarity coefficient) in GA3 E–I gene content between the first time point and every subsequent time point in adults and infants.

(C–E) Examples of E–I turnover events and corresponding strain replacement events are shown, including E–I replacement (C), T6SS loss (D), and T6SS gain (E). The plots in the upper and bottom left in each panel illustrate the estimated abundance of GA3 effector genes (measured as copies per *B. fragilis* genome) over time, with the plot in the upper right illustrating the estimated frequency of inferred strains in these samples. Only samples in which *B. fragilis* is present are shown. The bottom right plot illustrates the expected abundance of the various effector genes based on the effector genes encoded by reference strains that are phylogenetically close to the inferred strains.

See also Figure S4.

identified samples that contain *B. fragilis* but lack GA3-associated structural genes (STAR Methods) in both adult and infant datasets. Such samples indicate the presence of *B. fragilis* strains unable to intoxicate competitor bacteria using this pathway. We found that infant microbiomes containing *B. fragilis* are significantly less likely to lack GA3-associated structural genes relative to those of adults (Fisher's exact test, $p < 0.01$, 8% infants, 23% adults; n = 276; Figures 3A, S4A, and S4B).

This finding suggests that GA3 provides an advantage for *B. fragilis* in early life; however, the selective pressure underlying this advantage remained unclear. Several independent studies using gnotobiotic mice have shown that the GA3 T6SS can play a major role in the competition between *B. fragilis* strains in the gut (Chatzidaki-Livanis et al., 2016; Hecht et al., 2016; Wexler et al., 2016). However, *B. fragilis* is thought to be stable after acquisition from the mother, and inter-strain competition within the human gut microbiome has not been documented for this organism (Faith et al., 2013; Nayfach et al., 2016). Aiming to capture such processes in the developing microbiome, we estimated the abundance of GA3 E–I genes for individual infant samples as we did for adults. In general, the E–I landscape of infants mirrors that of adults, with generally a single genotype present in each sample. Moreover, many of the most prevalent E–I genotypes we observed in adults are also frequent in infants.

Notably, the infant microbiome datasets we analyzed include multiple samples per individual, thereby allowing us to examine the temporal dynamics of *B. fragilis* and of T6SS genes. Surprisingly, this analysis revealed many instances in which the E–I genotype of an individual changed between samples (Figure 3B). In total, we observed E–I turnover in 22 of the 117 infants for which longitudinal data were available. Such E–I turnover events include instances in which one GA3 genotype is replaced by another (Figure 3C), but also gains and losses of the T6SS (Figures 3D and 3E). To further confirm these E–I dynamics, we used the strain inference method described above. We detected a corresponding strain replacement in 17 of the 22 individuals in which an E–I turnover event was observed (Figures 3C–3E and S4C). Moreover, comparing the set of E–I genes detected in each sample to those encoded by the reference strains phylogenetically closest to the inferred strain, we further find overall agreement between observed and expected E–I turnover events. Notably, instances of one strain replaced by another with a similar E–I genotype (Figure S4C; Vatanen:T014827) or

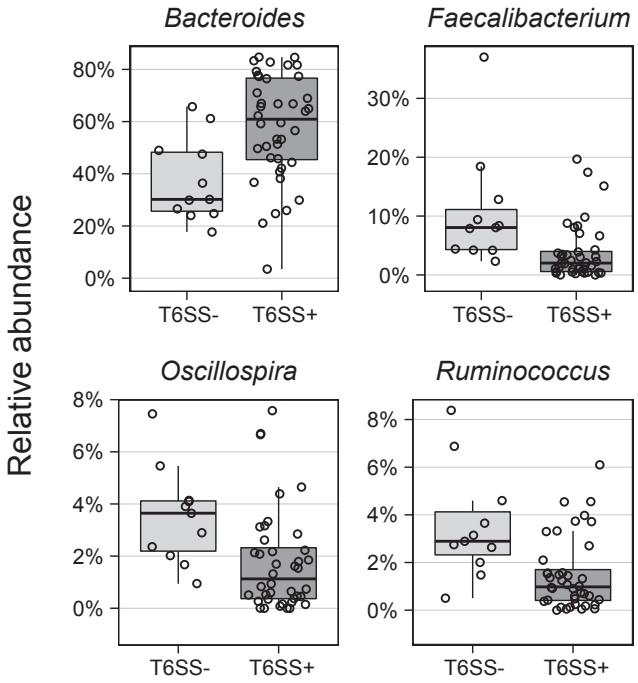


Figure 4. Differentially Abundant Genera between T6SS+ and T6SS– HMP Samples

Abundances are based on a 16S rRNA survey and only genera whose abundances are significantly different in T6SS+ versus T6SS− samples (at FDR < 0.05) are plotted. See also Table S1.

of transient co-existence of E–I genotypes (Figure S4C; Backhed:587) were also observed. Examination of the few HMP adult individuals for which data were available from multiple visits revealed one adult in which the E–I genotype similarly changed over time (Figure 3B). Similar analysis of GA1 and GA2 again revealed several individuals with non-conserved E–I profiles over time, suggesting a general instability of these subtypes in the developing gut microbiome (Figure S4D).

B. fragilis GA3 T6SS Is Associated with Shifts in Community Composition

Due to its lower frequency in adult microbiomes compared to those of infants, the GA3 T6SS is absent in many adult samples in which *B. fragilis* can be detected (23%), offering a unique opportunity to compare the community composition in samples with or without GA3. We hypothesized that such an analysis could identify potential competitors of *B. fragilis* targeted by the GA3 subtype. To this end, we obtained the taxonomic profile of all HMP samples (STAR Methods) and identified associations between these profiles and the presence of GA3 T6SS structural genes. We first compared overall community composition between samples as measured by the Bray-Curtis distance. We found that samples harboring *B. fragilis* and GA3 genes (T6SS+) significantly differ in community composition from samples harboring *B. fragilis* but lacking these genes (T6SS−; $p < 0.01$ PERMANOVA; $n = 51$). Examining the abundance of each genus across samples, we further identified four genera whose abundance in T6SS+ versus T6SS− samples significantly differs (Wilcoxon rank-sum test; false discovery rate [FDR] < 0.05; Figure 4; Table S1). Specifically, we found that the abundance of *Bacteroides* is positively correlated with the presence of GA3, which is consistent with experimental and theoretical work indicating that members of this genus are most likely to compete with *B. fragilis* for its niche (Trosvik and de Muinck, 2015). Furthermore, the genera *Faecalibacterium*, *Oscillospira*, and *Ruminococcus* from the phylum *Firmicutes* were negatively correlated with GA3. Gram-positive organisms are not targets of the T6SS; therefore, the observed decreases in abundance of these genera in T6SS+ microbiomes are likely to be the indirect result of selection for GA3 occurring in communities with an increased ratio of *Bacteroidetes* to *Firmicutes*.

DISCUSSION

Despite the wide distribution of T6S in Gram-negative bacteria, little is known about its role in natural communities. Here, we surveyed human microbiome samples and discovered that these communities are replete with bacteria containing the T6SS. We focused our analyses on genes specific to the order *Bacteroidales*; thus, our findings are an underestimate of the prevalence and impact of this system in gut communities. Nonetheless, our characterization of T6SS E–I gene distribution in human gut microbiomes suggests that this contact-dependent pathway plays a role in competition and selection at multiple levels.

We observed markedly low diversity of T6SS E–I genes in human microbiome samples. Specifically, a single genotype of GA1 and GA2 E–I genes is found in each microbiome, yet the abundance of these genes does not correlate with that of any one species in the *Bacteroides* genus. This supports a model in which antagonism via GA1 and GA2 exerts selective pressure for compatibility between *Bacteroides* spp. in the gut. We postulate that in the case of GA1 and GA2, horizontal transfer facilitates E–I compatibility. Indeed, theory predicts that strongly selected traits are most likely to be horizontally transferred in an environment in which ecological competition is strong, like the gut (Coyte et al., 2015; Niehus et al., 2015). Moreover, Comstock and colleagues found that transfer of GA1 and GA2 can occur between *Bacteroidales* species within the microbiome of an individual (Coyne et al., 2014).

Similar to GA1 and GA2, we found a single genotype of GA3 in each microbiome. However, GA3 is restricted to *B. fragilis* and our data are explained by the presence of a single *B. fragilis* strain in each individual. This is consistent with experimental studies demonstrating a role for GA3-dependent competition between *B. fragilis* strains (Chatzidaki-Livanis et al., 2016; Hecht et al., 2016; Wexler et al., 2016). Whether the infant microbiome strain replacements and accompanying E–I turnover events we observe are a consequence of GA3 activity cannot be determined from our current data. Nevertheless, these findings in conjunction with our observation that *B. fragilis* strains lacking the GA3 T6SS are more common in adults suggest that, early in life, *B. fragilis* strains compete for dominance. The infant microbiome may accordingly represent a particularly dynamic ecosystem in which the GA3 T6SS facilitates *B. fragilis* strain competition.

We find that strains of *B. fragilis* lacking GA3 are more commonly found in adults than infants. This could arise either

by the replacement of T6SS+ with T6SS− strains, or by the loss of the T6SS system from a previously T6SS+ strain of *B. fragilis*. This decline in *B. fragilis* GA3 prevalence in adulthood may reflect a change in its selective advantage. Indeed, there is precedent for the lability of T6S in bacteria undergoing strong shifts in environmental context, such as *Burkholderia mallei* and *Bordetella* spp. (Schwarz et al., 2010). It is likely that community effects buffering the *B. fragilis* niche develop with the maturation of the more stable adult gut community. Stabilization over development appears to render GA3 dispensable in certain contexts; for instance, within those microbiomes that contain lower populations of potential *B. fragilis* competitors and known targets of the GA3 pathway, other *Bacteroides* spp.

T6SS activity typically results in growth arrest or the lysis of competitor cells (Russell et al., 2014a). Thus, we anticipated that GA3-containing microbiomes would contain lower relative abundance of taxa antagonized by *B. fragilis* than those lacking GA3. Our counterintuitive finding that GA3 is enriched in microbiomes containing a large *Bacteroides* population suggests instead that *B. fragilis* primarily faces selective pressure from closely related species. Importantly, *B. fragilis* abundance is low compared to the *Bacteroides* consortium as a whole. It is therefore likely that the observed association between GA3 and community assembly reflects selection on GA3 mediated by community composition, rather than GA3-mediated impact on overall community assembly. While we do not detect other Gram-negative genera whose abundance is specifically lowered in GA3+ microbiomes, we cannot rule out that competition from less common or low abundance genera that fall below our detection limit might also select for the retention of GA3 in adults.

We showed here that a systematic characterization and large-scale computational analysis of metagenomic data can provide a means of linking the presence and abundance of T6 genes to microbial community composition. A caveat of our approach is that it does not account for differences in gene expression that could alter the phenotype elicited by the T6SS. Future work that integrates meta-transcriptomic data could provide a more sensitive measure of T6SS activity. For contact-dependent pathways like the T6SS, metagenomic analyses can provide a unique window into community biogeography. Indeed, *Bacteroides* spp. are thought to occupy a crowded niche proximal to the gut mucosa and our findings herein provide evidence of extensive cell-cell contacts between species of the genus (Whitaker et al., 2017). The T6SS is one of many antagonistic pathways whose operation is determined by the presence or absence of polymorphic toxins and corresponding antitoxins (Aoki et al., 2010; Whitney et al., 2017; Zhang et al., 2012). Thus, our study offers an analytical framework for more globally deciphering the forces that dictate the establishment and maintenance of bacterial communities.

STAR★METHODS

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

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Identifying T6SS Effector and Immunity Genes
  - Strain Sequencing
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Metagenomic and Genomic Data
  - Identifying Species-Specific Marker Genes
  - Estimating Gene Abundance in Microbiomes
  - Nucleotide Diversity Calculation
  - Inferring *B. fragilis* Strains
  - Phylogenetic Analysis
  - Predicting E–I Genes of Inferred *B. fragilis*
  - Identifying T6SS in Microbiome Samples
  - Community Composition Analysis
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.chom.2017.08.010.

AUTHOR CONTRIBUTIONS

B.D.R. and J.D.M. conceived the study. A.J.V., B.D.R., J.D.M., and E.B. designed the study. A.J.V. conducted the computational analysis. B.D.R., M.C.R., Y.B., and A.L.G. conducted experimental work. A.J.V., B.D.R., J.D.M., and E.B. wrote the paper.

ACKNOWLEDGMENTS

We thank Eric Alm and Chris Smillie for sharing StrainFinder code and for their support in running it, UW Genome Sciences ITS for high-performance computing resources, S. Brook Peterson for careful review of the manuscript, and Borenstein and Mougous lab members for helpful discussions. This work was supported by NIH grants GM118159 (A.L.G.) and AI080609 (J.D.M.), NIH New Innovator Award DP2AT00780201 (to E.B.), and the Pew Scholars Program (A.L.G.). A.J.V. was supported by a postdoctoral fellowship from the Natural Sciences and Engineering Research Council of Canada (NSERC). B.D.R. was supported by a Simons Foundation-sponsored Life Sciences Research Foundation postdoctoral fellowship. J.D.M. and A.L.G. hold Investigator in the Pathogenesis of Infectious Disease Awards from the Burroughs Wellcome Fund (BWF 1010010 and BWF 1014860), and J.D.M. is an HHMI investigator.

Received: May 18, 2017
Revised: July 7, 2017
Accepted: August 22, 2017
Published: September 13, 2017

REFERENCES

Anderson, M.C., Vonaesch, P., Saffarian, A., Marteyn, B.S., and Sansonetti, P.J. (2017). Shigella sonnei encodes a functional T6SS used for interbacterial competition and niche occupancy. Cell Host Microbe *21*, 769–776.e3.

Aoki, S.K., Diner, E.J., de Roodenbeke, C.T., Burgess, B.R., Poole, S.J., Braaten, B.A., Jones, A.M., Webb, J.S., Hayes, C.S., Cotter, P.A., and Low, D.A. (2010). A widespread family of polymorphic contact-dependent toxin delivery systems in bacteria. Nature *468*, 439–442.

Bäckhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P., Li, Y., Xia, Y., Xie, H., Zhong, H., et al. (2015). Dynamics and stabilization of the human gut microbiome during the first year of life. Cell Host Microbe *17*, 690–703.

Benz, J., and Meinhart, A. (2014). Antibacterial effector/immunity systems: it's just the tip of the iceberg. Curr. Opin. Microbiol. *17*, 1–10.

Bulgarelli, D., Garrido-Oter, R., Münch, P.C., Weiman, A., Dröge, J., Pan, Y., McHardy, A.C., and Schulze-Lefert, P. (2015). Structure and function of the

bacterial root microbiota in wild and domesticated barley. Cell Host Microbe *17*, 392–403.

Chatzidaki-Livanis, M., Geva-Zatorsky, N., and Comstock, L.E. (2016). Bacteroides fragilis type VI secretion systems use novel effector and immunity proteins to antagonize human gut Bacteroidales species. Proc. Natl. Acad. Sci. USA *113*, 3627–3632.

Coyne, M.J., Zitomersky, N.L., McGuire, A.M., Earl, A.M., and Comstock, L.E. (2014). Evidence of extensive DNA transfer between bacteroidales species within the human gut. MBio *5*, e01305–e01314.

Coyne, M.J., Roelofs, K.G., and Comstock, L.E. (2016). Type VI secretion systems of human gut Bacteroidales segregate into three genetic architectures, two of which are contained on mobile genetic elements. BMC Genomics *17*, 58.

Coyte, K.Z., Schluter, J., and Foster, K.R. (2015). The ecology of the microbiome: Networks, competition, and stability. Science *350*, 663–666.

Cullen, T.W., Schofield, W.B., Barry, N.A., Putnam, E.E., Rundell, E.A., Trent, M.S., Degnan, P.H., Booth, C.J., Yu, H., and Goodman, A.L. (2015). Gut microbiota. Antimicrobial peptide resistance mediates resilience of prominent gut commensals during inflammation. Science *347*, 170–175.

Delcher, A.L., Bratke, K.A., Powers, E.C., and Salzberg, S.L. (2007). Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics *23*, 673–679.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics *26*, 2460–2461.

Egan, F., Reen, F.J., and O'Gara, F. (2015). Tle distribution and diversity in metagenomic datasets reveal niche specialization. Environ. Microbiol. Rep. *7*, 194–203.

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

Finn, R.D., Clements, J., Arndt, W., Miller, B.L., Wheeler, T.J., Schreiber, F., Bateman, A., and Eddy, S.R. (2015). HMMER web server: 2015 update. Nucleic Acids Res. *43* (W1), W30–W38.

Hecht, A.L., Casterline, B.W., Earley, Z.M., Goo, Y.A., Goodlett, D.R., and Bubeck Wardenburg, J. (2016). Strain competition restricts colonization of an enteric pathogen and prevents colitis. EMBO Rep. *17*, 1281–1291.

Hibbing, M.E., Fuqua, C., Parsek, M.R., and Peterson, S.B. (2010). Bacterial competition: surviving and thriving in the microbial jungle. Nat. Rev. Microbiol. *8*, 15–25.

Hood, R.D., Singh, P., Hsu, F., Güvener, T., Carl, M.A., Trinidad, R.R., Silverman, J.M., Ohlson, B.B., Hicks, K.G., Plemel, R.L., et al. (2010). A type VI secretion system of Pseudomonas aeruginosa targets a toxin to bacteria. Cell Host Microbe *7*, 25–37.

Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. Nature *486*, 207–214.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647–1649.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. *10*, 845–858.

Kostic, A.D., Gevers, D., Siljander, H., Vatanen, T., Hyötyläinen, T., Hämäläinen, A.M., Peet, A., Tillmann, V., Pöhö, P., Mattila, I., et al.; DIABIMMUNE Study Group (2015). The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. Cell Host Microbe *17*, 260–273.

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

Levy, R., and Borenstein, E. (2013). Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. Proc. Natl. Acad. Sci. USA *110*, 12804–12809.

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

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

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

Niehus, R., Mitri, S., Fletcher, A.G., and Foster, K.R. (2015). Migration and horizontal gene transfer divide microbial genomes into multiple niches. Nat. Commun. *6*, 8924.

Prosser, J.I., Bohannan, B.J., Curtis, T.P., Ellis, R.J., Firestone, M.K., Freckleton, R.P., Green, J.L., Green, L.E., Killham, K., Lennon, J.J., et al. (2007). The role of ecological theory in microbial ecology. Nat. Rev. Microbiol. *5*, 384–392.

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

Rakoff-Nahoum, S., Foster, K.R., and Comstock, L.E. (2016). The evolution of cooperation within the gut microbiota. Nature *533*, 255–259.

Riley, M.A., and Wertz, J.E. (2002). Bacteriocins: evolution, ecology, and application. Annu. Rev. Microbiol. *56*, 117–137.

Russell, A.B., Peterson, S.B., and Mougous, J.D. (2014a). Type VI secretion system effectors: poisons with a purpose. Nat. Rev. Microbiol. *12*, 137–148.

Russell, A.B., Wexler, A.G., Harding, B.N., Whitney, J.C., Bohn, A.J., Goo, Y.A., Tran, B.Q., Barry, N.A., Zheng, H., Peterson, S.B., et al. (2014b). A type VI secretion-related pathway in Bacteroidetes mediates interbacterial antagonism. Cell Host Microbe *16*, 227–236.

Sana, T.G., Flaugnatti, N., Lugo, K.A., Lam, L.H., Jacobson, A., Baylot, V., Durand, E., Journet, L., Cascales, E., and Monack, D.M. (2016). Salmonella Typhimurium utilizes a T6SS-mediated antibacterial weapon to establish in the host gut. Proc. Natl. Acad. Sci. USA *113*, E5044–E5051.

Schwarz, S., Hood, R.D., and Mougous, J.D. (2010). What is type VI secretion doing in all those bugs? Trends Microbiol. *18*, 531–537.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313.

Trosvik, P., and de Muinck, E.J. (2015). Ecology of bacteria in the human gastrointestinal tract–identification of keystone and foundation taxa. Microbiome *3*, 44.

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

Vatanen, T., Kostic, A.D., d'Hennezel, E., Siljander, H., Franzosa, E.A., Yassour, M., Kolde, R., Vlamakis, H., Arthur, T.D., Hämäläinen, A.M., et al.; DIABIMMUNE Study Group (2016). Variation in microbiome LPS immunogenicity contributes to autoimmunity in humans. Cell *165*, 842–853.

Walter, J., and Ley, R. (2011). The human gut microbiome: ecology and recent evolutionary changes. Annu. Rev. Microbiol. *65*, 411–429.

Wexler, A.G., Bao, Y., Whitney, J.C., Bobay, L.M., Xavier, J.B., Schofield, W.B., Barry, N.A., Russell, A.B., Tran, B.Q., Goo, Y.A., et al. (2016). Human symbionts inject and neutralize antibacterial toxins to persist in the gut. Proc. Natl. Acad. Sci. USA *113*, 3639–3644.

Whitaker, W.R., Shepherd, E.S., and Sonnenburg, J.L. (2017). Tunable expression tools enable single-cell strain distinction in the gut microbiome. Cell *169*, 538–546.e12.

Whitney, J.C., Peterson, S.B., Kim, J., Pazos, M., Verster, A.J., Radey, M.C., Kulasekara, H.D., Ching, M.Q., Bullen, N.P., Bryant, D., et al. (2017). A broadly distributed toxin family mediates contact-dependent antagonism between gram-positive bacteria. Elife *6*, e26938.

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

Yassour, M., Vatanen, T., Siljander, H., Hämäläinen, A.M., Härkönen, T., Ryhänen, S.J., Franzosa, E.A., Vlamakis, H., Huttenhower, C., Gevers, D., et al.; DIABIMMUNE Study Group (2016). Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. Sci. Transl. Med. *8*, 343ra81.

Zhang, D., de Souza, R.F., Anantharaman, V., Iyer, L.M., and Aravind, L. (2012). Polymorphic toxin systems: Comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. Biol. Direct *7*, 18.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental Model and Subject Details		
Stool samples	Cullen et al., 2015	N/A
Deposited Data		
B. fragilis genomes from colony isolates	This paper	NCBI BioProject SRA: PRJNA375094
Oligonucleotides		
B. fragilis 16S F	This paper	ACGCTAGCTACAGGCTTAACACATGC
B. fragilis 16S R	This paper	GGACTACCAGGGTATCTAATCCTGTT
Bacteroides gyrB F	This paper	AGAAAACGCCCTGCGATGTACATTG
Bacteroides gyrB R	This paper	ATCGATATTTGCATACGTTGCATT
Software and Algorithms		
Bowtie 2	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
MetaPhlAn 2	Truong et al., 2015	http://huttenhower.sph.harvard.edu/metaphlan2
Usearch	Edgar, 2010	http://www.drive5.com/usearch/
Phyre	Kelley et al., 2015	http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index
Hmmer	Finn et al., 2015	http://hmmer.org/
samtools	Li et al., 2009	http://samtools.sourceforge.net/
StrainFinder	Alm Lab	https://github.com/cssmillie/StrainFinder
MAFFT	Katoh and Standley, 2013	http://mafft.cbrc.jp/alignment/software/
RAxML	Stamatakis, 2014	https://sco.h-its.org/exelixis/software.html
Geneious	Kearse et al., 2012	vR10.1.3
Glimmer	Delcher et al., 2007	http://www.cbcb.umd.edu/software/glimmer/index.shtml
Other		
Human microbiome project metagenomics data	Human Microbiome Project Consortium, 2012	https://www.hmpdacc.org/resources/data_browser.php
Metahit metagenomics data	Qin et al., 2010	EBI: ERA000116
Infant microbiome development metagenomics data	Vatanen et al., 2016	https://pubs.broadinstitute.org/diabimmune/three-country-cohort/resources/metagenomic-sequence-data
Infant microbiome development metagenomics data	Kostic et al., 2015	https://pubs.broadinstitute.org/diabimmune/t1d-cohort/resources/metagenomic-sequence-data
Infant microbiome development metagenomics data	Yassour et al., 2016	https://pubs.broadinstitute.org/diabimmune/antibiotics-cohort/resources/metagenomic-sequence-data
Infant microbiome development metagenomics data	Bäckhed et al., 2015	EBI: ERP005989

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Elhanan Borenstein (elbo@uw.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All human studies were conducted with the permission of the Yale Human Investigation Committee and informed consent was collected from all volunteers prior to participation. Recruitment of healthy human volunteers, sample collection, anaerobic processing, and −80°C storage of fecal samples in cryoprotectant under anaerobic conditions was previously described (Cullen et al., 2015). Briefly, the sample size of 30 donors (20 male and 10 female) was selected to represent common microbiome variation among humans. The donors were all greater than 18 years of age and were recruited without regard to sex. Samples were not assigned into groups. 16S rRNA gene sequencing was performed and used to identify samples which were most likely to harbor *Bacteroides fragilis*. The two samples chosen for strain isolation were a 24 year old male and a 54 year old female.

METHOD DETAILS

Identifying T6SS Effector and Immunity Genes

We sought to comprehensively catalog *Bacteroidales* T6SS E–I genes from reference genomes. In *Bacteroidales*, as in other bacteria, E–I genes are encoded adjacent to the genes for secreted structural proteins Hcp and VgrG. Accordingly, we manually curated genes adjacent to these structural genes across all publically available *Bacteroidales* genomes. Identified genes exhibited reduced GC content relative to the rest of the T6SS locus or the genome as a whole, and were encoded in bi-cistrons. Putative effectors always lacked characteristic signal peptides, consistent with transport via the T6SS apparatus, while putative immunity genes often encoded proteins with signal peptides. We used structural homology prediction (Phyre) (Kelley et al., 2015) and remote sequence homology search algorithms (Hmmer) (Finn et al., 2015) to predict functions for these genes, identifying many genes with functions associated with known T6SS toxin effectors. As in Coyne et al. (2016), our list included predicted cell-wall degrading enzymes, lipases, and nucleases as well as putative effector domains fused to either PAAR domains (DUF4280) or Hcp (Table S2). We additionally searched for contigs assembled from HMP metagenomes that contained subtype-specific T6SS structural gene sequences but lacked any E–I gene curated from reference genomes. Gene prediction was performed on these contigs using Glimmer within Geneious R10.1.3 and candidate genes were defined using the criteria as for reference genomes (Delcher et al., 2007; Kearse et al., 2012).

Strain Sequencing

Stool samples from four healthy individuals frozen in sterile glycerol (Cullen et al., 2015) were plated onto *Bacteroides* Bile Esculin or *Brucella* Blood Agar plates (BD Biosciences), to select for *Bacteroidales* colonies. Single colonies were picked into Mega Medium (Wu et al., 2015) and grown to stationary phase in anaerobic conditions at 37°C before freezing in 10% glycerol in 96-well plates. PCR was performed directly from the frozen glycerol stocks using primers to amplify the V1-V4 region of the 16S rRNA gene. PCR products were then Sanger sequenced. Sanger sequencing reads were converted to fastq format and NCBI Blast 2.2.31+ was used to align sequences to the SILVA 123 and GreenGenes 2011-1 16S rRNA gene databases in order to identify *Bacteroides fragilis*. To verify the *B. fragilis*-positive colonies, a second round of PCR was performed using primers to amplify and sequence the gyrB gene. Two of the four donors were confirmed to have *B. fragilis*. Twenty confirmed colonies from each *B. fragilis*-positive donor were then grown to stationary phase in TYG medium under anaerobic conditions at 37°C. Genomic DNA was isolated using the QIAGEN DNeasy Blood and Tissue Kit and prepared for whole genome sequencing using the MiSeq V3 Reagent Kit. Sequencing was performed in the Nickerson lab core facility in the UW Department of Genome Sciences. Sequencing reads were mapped to the set of *B. fragilis*-specific marker genes to generate alignments. Samples under 10x mean alignment read coverage were then discarded. Consensus sequence for each remaining sample was generated using the GATK FastaAlternateReferenceMaker. Subsequently, we constructed multi-alignments for all the samples using MAFFT 7.237, concatenated them, and then inferred a phylogenetic tree using the GTRGAMMAI model from RAxML 8.2.8 (Stamatakis, 2014).

QUANTIFICATION AND STATISTICAL ANALYSIS

Metagenomic and Genomic Data

Our analysis utilizes short read metagenomic data from several large-scale microbiome datasets. For adult microbiomes we downloaded 147 shotgun samples from HMP (Human Microbiome Project Consortium, 2012), and 99 healthy human shotgun samples from MetaHIT (Qin et al., 2010). Since an excessive fraction of human DNA will likely not markedly impact our ability to quantify *B. fragilis* abundance, HMP samples which failed QC were nonetheless included in our analysis. For infant microbiomes, we downloaded 300 samples from a study of development of the microbiome in the first year of life (Bäckhed et al., 2015), 769 samples from a study of autoimmune diseases (Vatanen et al., 2016), 237 samples from a study of antibiotic usage (Yassour et al., 2016), and 126 samples from a study of the development of Type 1 Diabetes (Kostic et al., 2015). Several of these datasets include multiple longitudinal samples from the same individuals, which were used for temporal analysis.

We downloaded all available *B. fragilis* genomes from RefSeq. Sequences from 3 strains were found to be contaminated with contigs matching species other than *B. fragilis*, and were discarded. A group of 8 strains appeared to be very distant in sequence homology from the rest of the strains, and were also discarded. We additionally downloaded from RefSeq all genomes of other *Bacteroides* species for which at least 5 strain genomes were available.

Identifying Species-Specific Marker Genes

We compiled a list of marker genes that could be used for strain-level inference. Our marker gene approach is similar to that used by MetaPhlAn (Truong et al., 2015), but relies on a more stringent selection of marker genes, supporting a more robust comparison at the strain level. Specifically, for our analysis, we identified a subset of the MetaPhlAn marker genes that are found in the genome of *every* sequenced strain in a single copy. To this end, for each of the MetaPhlAn marker genes associated with a given species, we used BLASTn to find every homolog (> 60% identity) in all the strains of that species. We used Usearch (Edgar, 2010) to cluster this larger set of genes into groups with > 90% identity and if a cluster with exactly one gene in each strain could be found, the marker gene was included in our list.

Estimating Gene Abundance in Microbiomes

We aligned shotgun reads single end using Bowtie2 (using parameters –a –N 1) to the set of genes of interest. Alignments with less than 97% identity, a quality score below 20, or multiple hits were discarded. To quantify the abundance of each gene, the number of reads aligned to this gene was normalized by the length of the gene and the total number of reads in the sample.

The average abundance of species-specific marker genes identified above was used as a proxy for the abundance of that species in the sample. We defined samples as having *B. fragilis* present if at least 100 reads could be aligned to *B. fragilis*-specific marker genes. When characterizing strain replacement, for which higher coverage of *B. fragilis* genes is required, we used instead a threshold of 500 reads. Because GA1 and GA2 are not restricted to a single species, and because *Bacteroidales* composition can vary dramatically, we considered samples with 100 reads mapping to the GA1 and GA2 E–I genes. GA1 and GA2 effector sequences contain repeats in the Rhs repeats which could interfere with read mapping, and therefore, we only used the toxin region (450bp from the C terminus) when quantifying GA1 or GA2 effector gene abundance. For GA3 we defined vgrG, TssP, TssN, TssK, TssB, TssO, TssG, TssF, and TssR as the Type VI structural genes, for GA1 we used TagB and for GA2 we used TagA. We selected these structural genes because based on reference genomes it was clear that a 97% identity similarity would unambiguously distinguish between subtypes.

Nucleotide Diversity Calculation

To estimate nucleotide diversity across species-specific marker genes, we again aligned all short reads in each sample to these genes. The obtained alignments were converted into a pileup using mpileup from samtools (parameters–excl-flags UNMAP,QCFAIL,DUP -A -q0 -C0 –B), and finally into an allele count matrix. The first and last 10 bases of each gene were discarded from the allele count matrix as we found they contained many poor quality alignments. We focused on high-coverage loci only, ignoring all loci where the coverage was less than 5X. If the number of high coverage sites was < 10% of the total length of the sequence, the sample was excluded from further consideration. Variable sites were defined as those having at least 2 counts of the minor allele. Nucleotide diversity was then calculated at these variant sites according to:

$$\pi = \frac{1}{n} \sum_i 2 p_i q_i$$

where n is the total length of the genes, i corresponds to the variable sites, and p and q correspond to the frequency of the major and minor allele at site i.

Inferring *B. fragilis* Strains

To infer strain diversity in each sample, we used StrainFinder (https://github.com/cssmillie/fmt), a previously introduced method for inferring the most likely set of strains in metagenomic samples based on nucleotide variants. To this end, we again aligned the short read in each sample to the set of *B. fragilis*-specific marker genes identified above, and converted the alignment to a count matrix describing the number of counts of each nucleotide at every position along the genes. As when calculating nucleotide diversity, we discarded the first and last 10 bases and only considered sites with at least 5 counts. We also discarded samples where the high coverage sites were less than 10% of the total length as they resulted in poor quality trees. When running StrainFinder we reduced the data to only those sites with population variability, defined, as above, as sites with at least 2 counts of the minor allele. As noted above, we only considered samples with a sufficient coverage on the *B. fragilis* marker genes to enable robust strain inference.

StrainFinder determines the relative strain abundance and genotypes at variable sites by considering the likelihood of the observed allele counts and using an expectation maximization approach. An optimal number of strains between 1 and 10 was determined using AIC. For each run of StrainFinder we used 5 independent runs of 200 expectation maximization iterations and selected the best fit; these parameters yield reproducible inference of strains. When analyzing temporal data with StrainFinder, we combined allele counts from all samples of an individual into a single 3-dimensional matrix. Using the genotypes from StrainFinder we then reconstructed strain-specific versions of each marker gene and then created subsequences consisting of only the high coverage sites we used in the analysis. Inferred strains were then further examined using a phylogenetic analysis as described below.

Phylogenetic Analysis

A phylogenetic tree of the reference *B. fragilis* strains was constructed based on their species-specific marker genes. Specifically, we aligned the strains' versions of each marker gene using MAFFT, concatenated the alignments of all genes, and then constructed a tree using the GTRGAMMAI model from RAxML (Stamatakis, 2014), as has been done previously (Wexler et al., 2016).

To determine whether the strains inferred by StrainFinder are monophlyletic, we combined the sequences from the inferred strains (as determined by StrainFinder), with the sequences from the available reference *B. fragilis* genomes, and recreated the strain phylogeny using the same method as described above. We defined inferred strains as monophyletic if their common ancestor does not have any descendent outside the set of inferred strains or if the distance was less than 0.001 substitutions per site.

Predicting E–I Genes of Inferred *B. fragilis*

We predicted the E–I gene content of an inferred *B. fragilis* strain by examining the E–I content in the genome of its nearest neighbors on a phylogenetic tree that contains the inferred strains from a given sample and the reference strains (as described above). Specifically, for every strain identified from StrainFinder, we identified the most recent ancestor that have both the inferred strain and at least one reference strain as descendants. We then used the average E–I content of all reference strains descendant from this ancestor as the predicted E–I content of the inferred strain. To then estimate the predicted E–I content in the sample, we combined the predicted E–I content of each inferred strain weighted by their relative abundance. To determine the confidence of the predicted E–I content we determined the average phylogenetic distance of these reference strains to the ancestor identified above.

Identifying T6SS in Microbiome Samples

For every sample, we estimated the number of reads expected to map to the *B. fragilis* GA3 T6SS structural genes based on the number of reads mapped to *B. fragilis*-specific marker genes in that sample and the ratio between the total length of *B. fragilis*-specific marker genes and *B. fragilis* T6SS structural genes. We define samples to be T6SS+ if *B. fragilis* was present (as defined above) and the number of reads mapped to T6SS structural genes was more than 10% of the expected number (and see Figure S4A). We define samples to be T6SS- if *B. fragilis* was present and the number of reads mapped to T6SS structural genes was less than 10% of the expected number.

Community Composition Analysis

To obtained independent estimate community composition in each sample, we downloaded the v35 16S OTU abundance table for human gut microbiomes from HMP (ftp://public-ftp.hmpdacc.org/HMQCP/otu_table_psn_v35.txt.gz), summed the counts from all OTUs in the same genus, and calculated the relative abundance of each genus. Importantly, because 16S sequencing depth is independent from the depth of shotgun samples used to determine T6SS+ versus T6SS- classification, using these 16S-based data allows us to compare T6SS presence with community taxonomic profiles without potential coverage-related biases. Samples were classified into T6SS+ and T6SS- as described above. GA1 and GA2 lack uniquely identifying structural genes so we defined T6SS+ versus T6SS- as samples with versus without 100 counts mapping to the GA1 or GA2 E–I genes respectively. The distance between samples was defined by the Bray-Curtis distance at the genus level, and significance of separation between T6SS+ and T6SS- samples was evaluated using PERMANOVA. For the subset of genera whose average abundance across samples was > 0.1%, we used a Wilcoxon rank sum test to compare their abundance in T6SS+ versus T6SS- samples using a 5% FDR.

DATA AND SOFTWARE AVAILABILITY

All sequences were deposited into NCBI SRA under BioProject ID SRA: PRJNA375094.

Cell Host & Microbe, Volume *22*

Supplemental Information

The Landscape of Type VI Secretion across Human Gut Microbiomes Reveals Its Role in Community Composition

Adrian J. Verster, Benjamin D. Ross, Matthew C. Radey, Yiqiao Bao, Andrew L. Goodman, Joseph D. Mougous, and Elhanan Borenstein

Figure S1

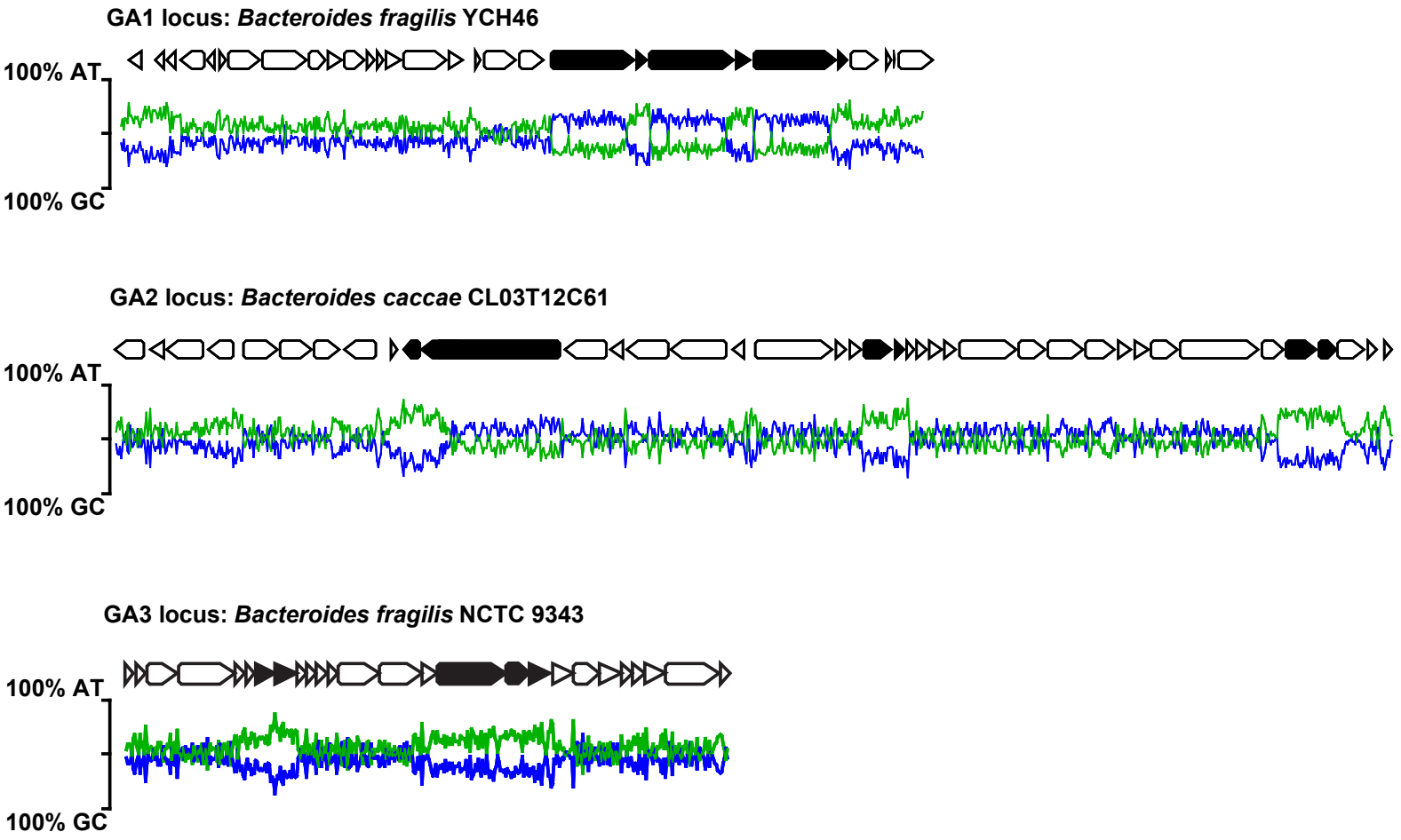


Figure S1. Genetic organization of the GA1-3 loci in Bacteroidales. Related to Figure 1.
Schematic of the GA1, GA2, and GA3 loci. E–I genes are filled in with black, while structural genes are outlined. GC and AT content is plotted below the gene diagrams for each locus.

Figure S2

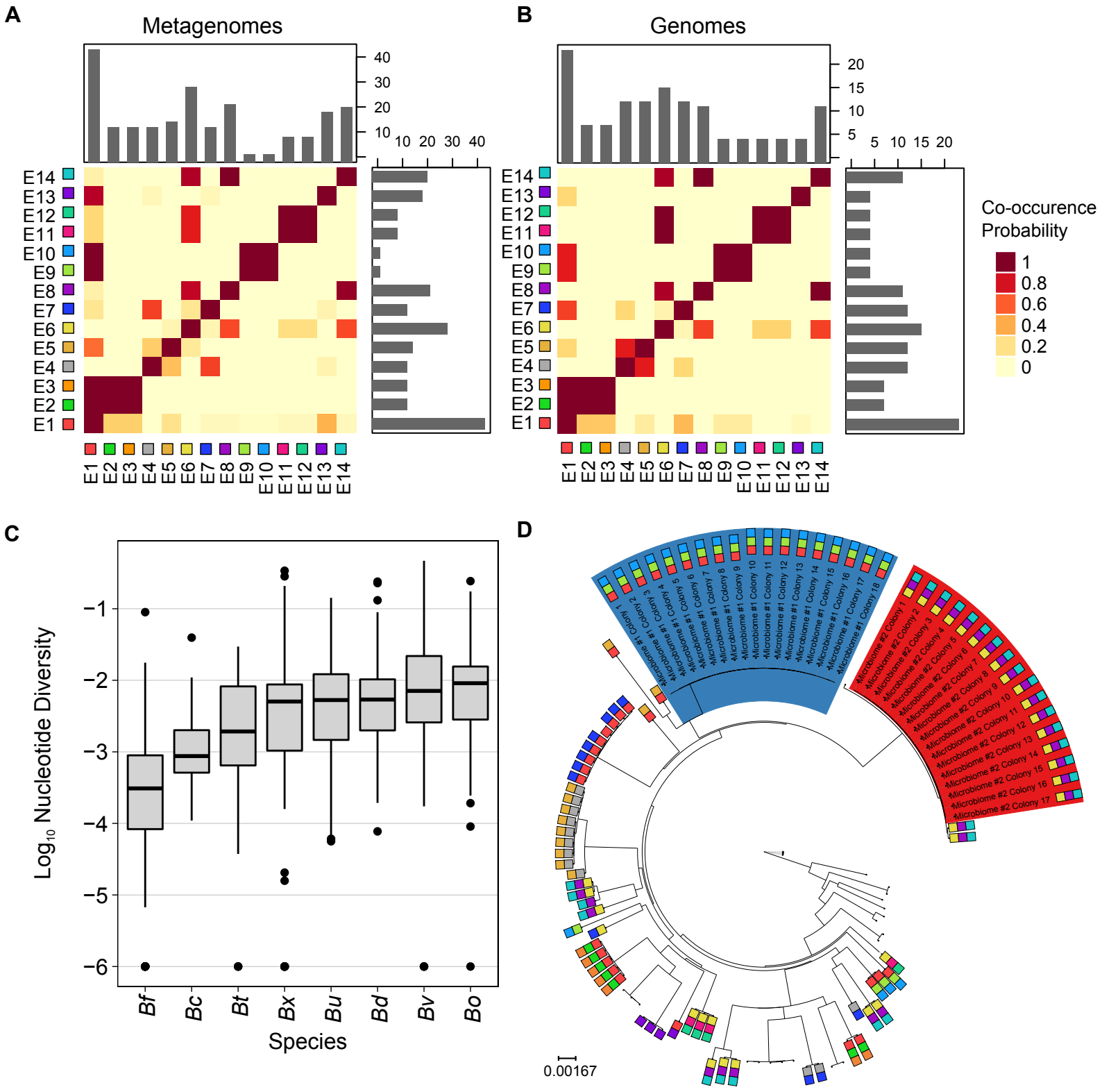


Figure S2. The co-occurrence of GA3 effector genes, and *Bacteroides* nucleotide and strain diversity in the gut microbiome of adults. Related to Figure 2.

(A-B) Each cell in the heatmap, a_{ij}, denotes the probability that effector gene *i* is detected in a metagenomic sample **(A)** or encoded in a genome **(B)** given that effector gene *j* was detected/encoded. The barplots on the top and right of each heatmap illustrate the number of metagenomes or genomes in which each effector gene was detected. **(C)** The nucleotide diversity for different Bacteroides spp. in adult samples from the HMP and MetaHIT studies. Nucleotide diversity was calculated based on population variants in species-specific marker genes (Experimental Procedures). Only species with at least 5 genomes in RefSeq were considered. *Bf:B. fragilis; Bc: B. cellulosilyticus; Bt: B. thetaiotaomicron; Bx: B. xylanisolvens; Bu: B. uniformis; Bd: B. dorei; Bv: B. vulgatus; Bo: B. ovatus*. **(D)** A phylogenetic tree linking previously sequenced *B. fragilis* reference genomes with sequenced colonies from two individuals (in red and blue). The effector genes encoded by each reference genome and the new sequenced genomes from stool are represented by colored squares as in Figure 1A.

Figure S3

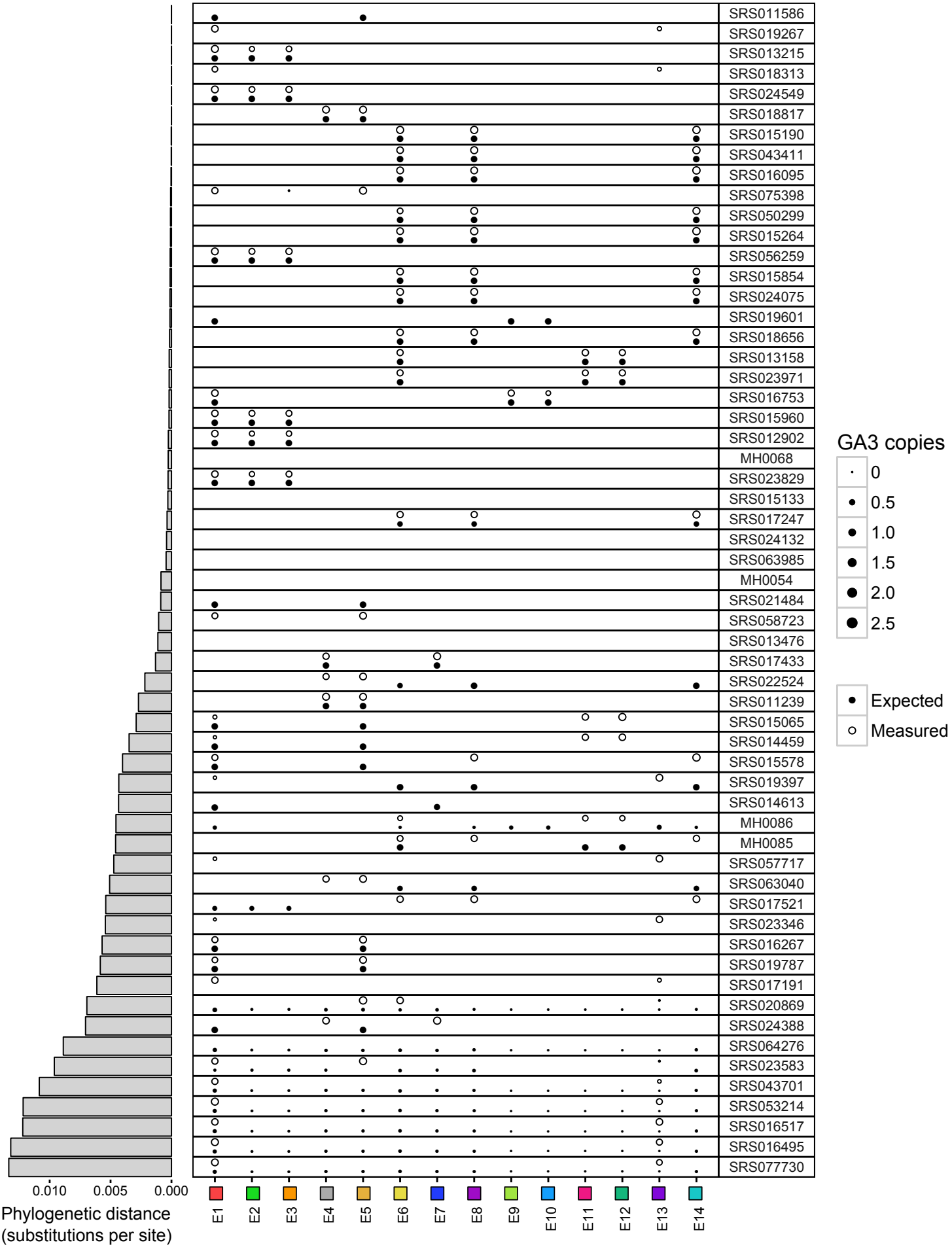


Figure S3. Measured and expected E–I gene abundances in HMP and MetaHIT samples. Related to Figure 2. Measured abundances (hollow circles) are based on short read mapping to effector genes. Expected abundances (filled circles) are based on the reference strains phylogenetically closest to the inferred strain. Point size is scaled to the calculated copy number of each effector. The barplot on the left shows the phylogenetic distance between the inferred strain and its nearest reference strain in the phylogenetic tree.

Figure S4

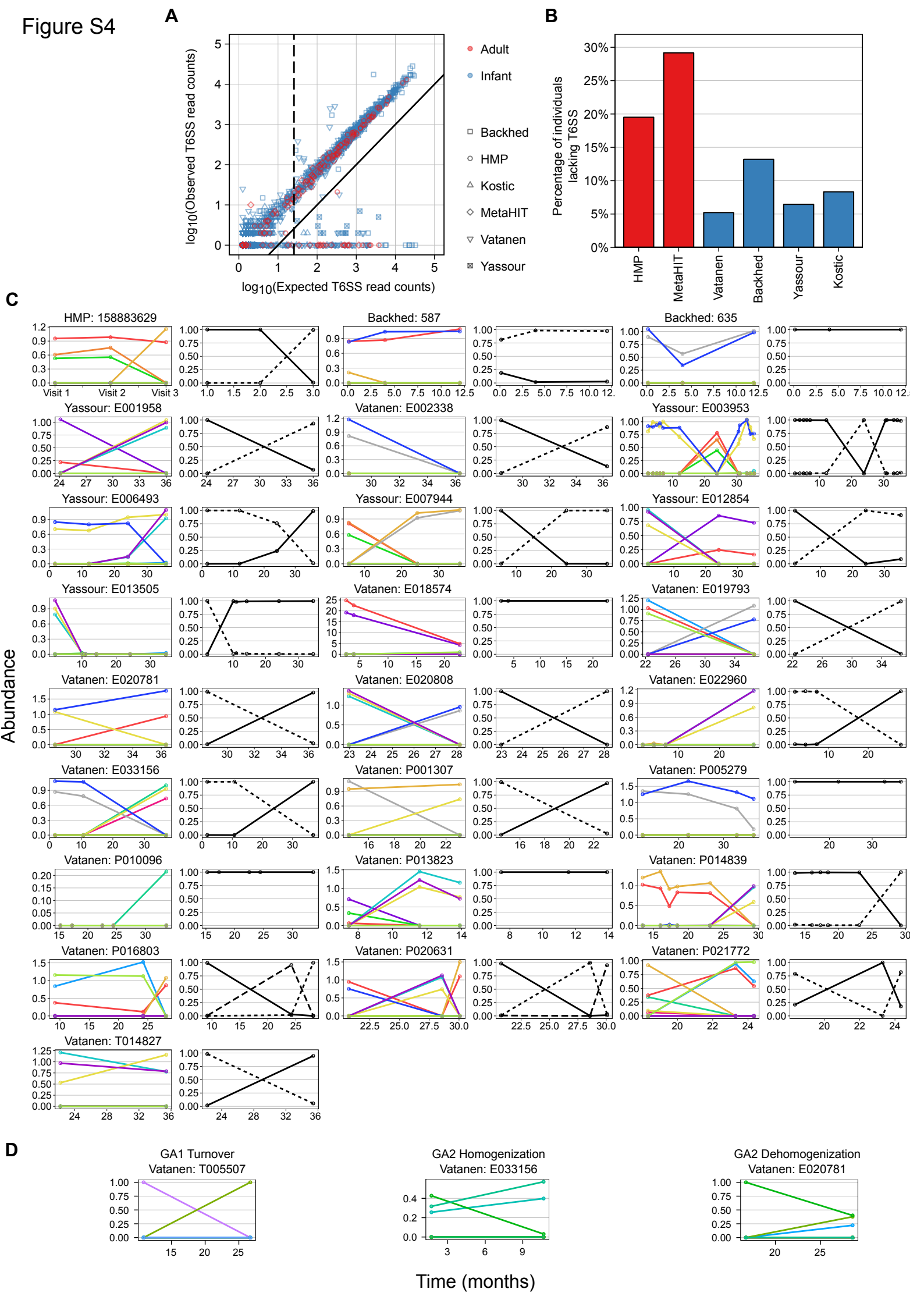

A
log10(Observed T6SS read counts)
log10(Expected T6SS read counts)
Adult
Infant
Backhed
HMP
Kostic
MetaHIT
Vatanen
Yassour
B
Percentage of individuals lacking T6SS
HMP
MetaHIT
Vatanen
Backhed
Yassour
Kostic
C
Abundance
HMP: 158883629
Visit 1
Visit 2
Visit 3
Backhed: 587
Backhed: 635
Yassour: E001958
Vatanen: E002338
Yassour: E003953
Yassour: E006493
Yassour: E007944
Yassour: E012854
Yassour: E013505
Vatanen: E018574
Vatanen: E019793
Vatanen: E020781
Vatanen: E020808
Vatanen: E022960
Vatanen: E033156
Vatanen: P001307
Vatanen: P005279
Vatanen: P010096
Vatanen: P013823
Vatanen: P014839
Vatanen: P016803
Vatanen: P020631
Vatanen: P021772
Vatanen: T014827
D
GA1 Turnover
Vatanen: T005507
GA2 Homogenization
Vatanen: E033156
GA2 Dehomogenization
Vatanen: E020781
Time (months)

Figure S4. The prevalence of T6SS- samples and patterns of E-I turnover. Related to Figure 3.
(A) The prevalence of T6SS- samples in infants and adults as ascertained through the relationship between the number of reads expected and measured to map to the GA3 T6SS structural genes, from each adult (red) and infant (blue) gut microbiome sample. The expected number is based on the number of reads mapping to *B. fragilis*-specific marker genes, normalized by gene lengths. The dotted line represents the cutoff used for determining that *B. fragilis* is present in a sample. The solid line represents an observed number of reads that is 10% of expected, and was used to distinguish T6SS+ from T6SS- samples. As evident by this plot, T6SS+ and T6SS- samples can be clearly defined. Different shapes correspond to different datasets, with adult samples colored in red and infant samples in blue. **(B)** The percentage of individuals of those harboring B. fragilis, that lack the GA3 T6SS across different adult (red) and infant (blue) datasets. Individuals that were not consistent in terms of T6SS+/- classification across different time points were not included. **(C)** E–I turnover and strain replacement for GA3. Details and format are as in Figure 3C-E (top plots). Each pair of plots is labeled with the individual code and the dataset it comes from. **(D)** Analysis of GA1 and GA2 over time has demonstrated similar patterns to those observed for GA3, with most individuals exhibiting a conserved effector profile over time. Among the few individuals that exhibited different E-I profiles at different time points, we can observe examples of E-I turnover, E–I genotype homogenization (a decrease in diversity over time), and E-I genotype dehomogenization. Here the abundance on the y-axis refers to the relative abundance of GA1 or GA2 effector genes.

Table S1. Genus-level differential abundance in T6SS+ vs. T6SS- samples. Related to Figure 4.
Abundances are based on a 16S rRNA survey and only genera which are sufficiently abundant across T6SS+ vs. T6SS- samples are listed.

Genus	Median Abundance T6SS+ (%)	Median Abundance T6SS- (%)	p-value	FDR Adjusted p-value	Passes FDR
Faecalibacterium	3.8	10.6	8.30E-04	1.70E-02	TRUE
Bacteroides	57.9	37.5	3.10E-03	1.90E-02	TRUE
Ruminococcus	1.5	3.5	3.10E-03	1.90E-02	TRUE
Oscillospira	1.8	3.5	3.60E-03	1.90E-02	TRUE
Eubacterium	0.3	0.4	1.70E-02	7.20E-02	FALSE
Odoribacter	0.4	0.8	2.80E-02	9.80E-02	FALSE
Subdoligranulum	1.9	3.3	3.50E-02	1.10E-01	FALSE
Sutterella	2.4	0.1	7.60E-02	2.00E-01	FALSE
Dialister	1.2	0.7	1.10E-01	2.50E-01	FALSE
Clostridium	1.5	0.9	1.60E-01	3.20E-01	FALSE
Alistipes	5.7	6.7	1.80E-01	3.20E-01	FALSE
Coprococcus	0.5	0.6	1.80E-01	3.20E-01	FALSE
Akkermansia	0.1	0.3	2.20E-01	3.50E-01	FALSE
Lachnospira	0.7	0.7	2.60E-01	3.90E-01	FALSE
Parabacteroides	4.5	2.1	3.00E-01	4.20E-01	FALSE
Roseburia	1.5	2.7	4.50E-01	5.70E-01	FALSE
Blautia	0.9	0.5	4.60E-01	5.70E-01	FALSE
Megamonas	0.2	0	5.30E-01	6.20E-01	FALSE
Prevotella	0.1	0.3	6.00E-01	6.60E-01	FALSE
Phascolarctobacterium	0.8	0.8	8.00E-01	8.40E-01	FALSE
Escherichia	0.1	0.1	8.50E-01	8.50E-01	FALSE

Table S2. E–I gene catalog and reference strains used. Related to STAR Methods.

T6SS E–I accession numbers and locus tags are listed by genetic architecture subtype.
Reference strains used the study are listed in a separate sheet.

GA1

Effector #	Genome	Effector locus tag	Effector accession	Immunity #	Immunity locus tag	Immunity accesion
GA1_E1	*Bacteroides cellulosilyticus* CL02T12C19	HMPREF1062_03920	WP_007857070.1	GA1_I1	HMPREF1062_03918	WP_004323409.1
GA1_E2	*Parabacteroides* sp. D25	HMPREF0999_RS10520	WP_008669247.1	GA1_I2	HMPREF0999_RS10515	WP_008669248.1
GA1_E3	*Bacteroides coprophilus* DSM 18228	BACCOPRO_01421	WP_008141979.1	GA1_I3	BACCOPRO_01420	WP_008141977.1
GA1_E4	*Bacteroides fragilis* YCH46	BF2847	YP_100131.1	GA1_I4	BF2849	YP_100132.1
GA1_E5	*Bacteroides fragilis* YCH46	BF2850	YP_100133.1	GA1_I5	BF2851	YP_100134.1
GA1_E6	*Bacteroides fragilis* YCH46	BF2852	YP_100135.1	GA1_I6	BF2853	YP_100136.1
GA1_E7	*Bacteroides* sp. 3_2_5	BSHG_RS0220345	WP_007857070.1	GA1_I7	BSHG_4301	WP_008659502.1
GA1_E8	*Bacteroides uniformis* CL03T00C23	HMPREF1072_RS17450	WP_005836806.1	GA1_I8	HMPREF1072_RS17445	WP_005836803.1
GA1_E9	*Bacteroides vulgatus* 3775 SL(B) 10 (iv)	M097_RS18140	WP_011203027.1	GA1_I9	M097_RS18145	WP_008776200.1

GA2

Effector #	Genome	Effector locus tag	Effector accession	Immunity #	Immunity locus tag	Immunity accession
GA2_E1	*Parabacteroides distasonis* CL03T12C09	HMPREF1075_RS12295	WP_036629059.1	GA2_I1	HMPREF1075_RS12300	WP_005858631.1
GA2_E2	*Bacteroides fragilis* CL05T00C42	HMPREF1079_RS08215	WP_005802593.1	GA2_I2	HMPREF1079_RS08220	WP_005802592.1
GA2_E3	*Bacteroides eggerthi* DSM 20697	BACEGG_RS01920	WP_004288712.1	GA2_I3	BACEGG_RS01915	WP_039953011.1
GA2_E4	*Bacteroides caccae* CL03T12C61	HMPREF1061_RS00705	WP_005682920.1	GA2_I4	HMPREF1061_RS00700	WP_005682921.1
GA2_E5	*Bacteroides caccae* CL03T12C61	HMPREF1061_RS00780	WP_005682903.1	GA2_I5	HMPREF1061_RS00775	WP_005682904.1
GA2_E6	*Bacteroides vulgatus* dnLKV7	C800_RS02355	WP_016271567.1	GA2_I6	C800_RS02360	WP_016271566.1
GA2_E7	*Bacteroides vulgatus* dnLKV7	C800_RS02310	WP_016271575.1	GA2_I7	C800_RS02305	WP_016271576.1
GA2_E8	*Bacteroides dorei* DSM17855	BACDOR_RS16910	WP_007832792.1	GA2_I8	BACDOR_RS16905	WP_007832794.1
GA2_E9	*Parabacteroides* sp. 20_3	HMPREF9008_RS04985	WP_007832792.1	GA2_I9	HMPREF9008_RS04990	WP_005858575.1
GA2_E10	*Parabacteroides* sp. 20_3	HMPREF9008_RS04905	WP_008778001.1	GA2_I10	HMPREF9008_RS04900	WP_008778000.1
GA2_E11	*Parabacteroides* sp. 20_3	HMPREF9008_RS23040	WP_008778764.1	GA2_I11	HMPREF9008_RS23045	WP_005942723.1
GA2_E12	*Parabacteroides* sp. 20_3	HMPREF9008_RS22995	WP_005942780.1	GA2_I12	HMPREF9008_RS22990	WP_005942781.1
GA2_E13	*Bacteroides vulgatus* dnLKV7	C800_RS02240	WP_016271588.1	GA2_I13	C800_RS02360	WP_016271589.1
GA2_E14	*Bacteroides dorei* DSM17855	BACDOR_RS22955	WP_007832756.1	GA2_I14	BACDOR_RS17020	WP_032935744.1
GA2_E15	*Parabacteroides distasonis* CL03T12C09	HMPREF1075_RS12175	WP_007832792.1	GA2_I15	HMPREF1075_RS12170	WP_005858587.1
GA2_E16	*Bacteroides caccae* CL03T12C61	HMPREF1061_RS00825	WP_005682893.1	GA2_I16	HMPREF1061_RS00830	WP_050554721.1
GA2_E17	*Bacteroides dorei* DSM17855	BACDOR_RS16955	WP_007832774.1	GA2_I17	BACDOR_RS16960	WP_032935753.1
GA2_E18	*Bacteroides fragilis* CL05T00C42	HMPREF1079_RS08170	WP_005802603.1	GA2_I18	HMPREF1079_RS08165	WP_005802605.1

GA3

Effector #	Genome	Effector locus tag	Effector accession	Immunity #	Immunity locus tag	Immunity accession
GA3_E1	*Bacteroides fragilis* 3774_T13	M117_RS11205	WP_008659774.1	GA3_I1	M117_RS0122600	WP_008659772.1
GA3_E2	*Bacteroides fragilis* 3774_T13	M117_RS11240	WP_008659750.1	GA3_I2	M117_RS11245	WP_008659758.1
GA3_E3	*Bacteroides fragilis* 3774_T13	M117_RS23385	WP_050444646.1	GA3_I3	M117_RS11255	WP_008659747.1
GA3_E4	*Bacteroides fragilis* DS-166	M074_RS10745	WP_025814153.1	GA3_I4	M074_RS10750	WP_025814152.1
GA3_E5	*Bacteroides fragilis* DS-166	M074_RS10800	WP_011202656.1	GA3_I5	M074_RS10805	WP_011202655.1
GA3_E6	*Bacteroides fragilis* NCTC 9343	BF9343_1937	WP_010992803.1	GA3_I6	BF9343_1936	WP_005787090.1
GA3_E7	*Bacteroides fragilis* NCTC 9343	BF9343_1928	WP_010992798.1	GA3_I7	BF9343_1927	WP_010992797.1
GA3_E8	*Bacteroides fragilis* J38-1	M068_RS11120	WP_032561265.1	GA3_I8	M068_2000	WP_005787075.1
GA3_E9	*Bacteroides fragilis* Korea_419	M065_RS13210	WP_032600880.1	GA3_I9	M065_2764	WP_005794673.1
GA3_E10	*Bacteroides fragilis* Korea_419	M065_RS13185	WP_005794682.1	GA3_I10	M065_2759	WP_005794684.1
GA3_E11	*Bacteroides fragilis* B1_(UDC16-1)	M069_RS16890	WP_009292151.1	GA3_I11	M069_2106	WP_009292150.1
GA3_E12	*Bacteroides fragilis* B1_(UDC16-1)	M069_RS31145	WP_050446922.1	GA3_I12	M069_2103	WP_032571500.1
GA3_E13	*Bacteroides fragilis* DS-71	M073_RS11420	WP_032589942.1	GA3_I13	M073_1832	WP_032562840.1
GA3_E14	*Bacteroides fragilis* HMW616	HMPREF1205_RS15360	WP_005819888.1	GA3_I14	HMPREF1205_03575	WP_005819890.1

Genome	Strain	Assembly	GenBank FTP
Bacteroides cellulosilyticus	WH2	GCA_001318345.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_001318345.1_ASM131834v1
Bacteroides cellulosilyticus	DSM 14838	GCA_000158035.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000158035.1_ASM15803v1
Bacteroides cellulosilyticus	CL02T12C19	GCA_000273015.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000273015.1_Bact_cell_CL02T12C19_V1
Bacteroides cellulosilyticus	WH2	GCA_000463315.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000463315.1_B_cell_WH2_1.0
Bacteroides cellulosilyticus	CL09T06C25	GCA_001535595.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_001535595.1_ASM153559v1
Bacteroides dorei	DSM 17855	GCA_000156075.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000156075.1_ASM15607v1
Bacteroides dorei	5_1_36/D4	GCA_000158335.2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000158335.2_Bact_dorei_5_1_36_D4_V2
Bacteroides dorei	CL02T00C15	GCA_000273035.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000273035.1_Bact_dore_CL02T00C15_V1
Bacteroides dorei	CL02T12C06	GCA_000273055.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000273055.1_Bact_dore_CL02T12C06_V1
Bacteroides dorei	CL03T12C01	GCA_000273075.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000273075.1_Bact_dore_CL03T12C01_V1
Bacteroides dorei		GCA_000738045.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000738045.1_ASM73804v1
Bacteroides dorei		GCA_000738065.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000738065.1_ASM73806v1
Bacteroides dorei		GCA_001274835.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_001274835.1_ASM127483v1
Bacteroides ovatus	ATCC 8483	GCA_001314995.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_001314995.1_ASM131499v1
Bacteroides ovatus	ATCC 8483	GCA_000154125.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000154125.1_ASM15412v1
Bacteroides ovatus	3_8_47FAA	GCA_000218325.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000218325.1_Bact_ovat_3_8_47FAA_V1
Bacteroides ovatus	CL02T12C04	GCA_000273195.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000273195.1_Bact_ovat_CL02T12C04_V1
Bacteroides ovatus	CL03T12C18	GCA_000273215.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000273215.1_Bact_ovat_CL03T12C18_V1
Bacteroides ovatus	KLE1656	GCA_001578575.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_001578575.1_ASM157857v1
Bacteroides ovatus	SD CMC 3f	GCA_000178275.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000178275.1_ASM17827v1
Bacteroides ovatus	3725 D9 iii	GCA_000699665.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000699665.1_ASM69966v1
Bacteroides ovatus	3725 D1 iv	GCA_000699725.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000699725.1_ASM69972v1
Bacteroides ovatus	CL09T03C03	GCA_001535615.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_001535615.1_ASM153561v1
Bacteroides thetaiotaomicron	VPI-5482	GCA_000011065.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000011065.1_ASM1106v1
Bacteroides thetaiotaomicron	7330	GCA_001314975.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_001314975.1_ASM131497v1
Bacteroides thetaiotaomicron	dnLKV9	GCA_000403155.2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000403155.2_Bact_thet_dnLKV9_V1
Bacteroides thetaiotaomicron	KLE1254	GCA_001578565.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_001578565.1_ASM157856v1
Bacteroides thetaiotaomicron	2e6A	GCA_001373135.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_001373135.1_2e6A_assembly
Bacteroides thetaiotaomicron	19_BTHE	GCA_001055755.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_001055755.1_ASM105575v1
Bacteroides thetaiotaomicron	3a5B	GCA_000937835.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000937835.1_3a5B_assembly
Bacteroides thetaiotaomicron		GCA_001049535.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_001049535.1_3731
Bacteroides thetaiotaomicron		GCA_001049555.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_001049555.1_7330
Bacteroides uniformis	CL03T00C23	GCA_000273785.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000273785.1_Bact_unif_CL03T00C23_V1
Bacteroides uniformis	ATCC 8492	GCA_000154205.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000154205.1_ASM15420v1
Bacteroides uniformis	CL03T12C37	GCA_000273275.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000273275.1_Bact_unif_CL03T12C37_V1
Bacteroides uniformis	dnLKV2	GCA_000403175.2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000403175.2_Bact_unif_dnLKV2_V1
Bacteroides uniformis	KLE1607	GCA_001578555.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_001578555.1_ASM157855v1
Bacteroides uniformis	3978 T3 ii	GCA_000699825.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000699825.1_ASM69982v1
Bacteroides uniformis	3978 T3 i	GCA_000699885.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000699885.1_ASM69988v1
Bacteroides uniformis	2a7B	GCA_000939155.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000939155.1_2a7B_assembly3
Bacteroides uniformis	2a3A	GCA_001403955.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_001403955.1_2a7B_assembly2
Bacteroides vulgatus	ATCC 8482	GCA_000012825.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000012825.1_ASM1282v1
Bacteroides vulgatus	mpk	GCA_001412315.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_001412315.1_ASM141231v1
Bacteroides vulgatus	CL09T03C04	GCA_000273295.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000273295.1_Bact_vulg_CL09T03C04_V1
Bacteroides vulgatus	dnLKV7	GCA_000403235.2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000403235.2_Bact_vulg_dnLKV7_V1
Bacteroides vulgatus	PC510	GCA_000178195.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000178195.1_ASM17819v1
Bacteroides vulgatus	3775 SL(B) 10 (iv)	GCA_000699705.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000699705.1_ASM69970v1
Bacteroides vulgatus	3775 SR(B) 19	GCA_000699845.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000699845.1_ASM69984v1
Bacteroides vulgatus	3975 RP4	GCA_000699865.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000699865.1_ASM69986v1
Bacteroides xylanisolvens	CL03T12C04	GCA_000273315.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000273315.1_Bact_xyla_CL03T12C04_V1
Bacteroides xylanisolvens	SD CC 1b	GCA_000178215.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000178215.1_ASM17821v1
Bacteroides xylanisolvens	SD CC 2a	GCA_000178295.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000178295.1_ASM17829v1
Bacteroides xylanisolvens		GCA_000577295.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000577295.1_SD_CC_1b
Bacteroides xylanisolvens		GCA_000577955.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000577955.1_SD_CC_2a
Bacteroides xylanisolvens	XB1A	GCA_000210075.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000210075.1_ASM21007v1
Bacteroides fragilis	YCH46	GCA_000009925.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000009925.1_ASM992v1
Bacteroides fragilis	638R	GCA_000210835.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000210835.1_ASM21083v1
Bacteroides fragilis	BOB25	GCA_000965785.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000965785.1_ASM96578v1
Bacteroides fragilis	BE1	GCA_001286525.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_001286525.1_BFBE1.1
Bacteroides fragilis	CL07T00C01	GCA_000263115.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000263115.1_Bact_frag_CL07T00C01_V1
Bacteroides fragilis	CL03T00C08	GCA_000273095.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000273095.1_Bact_frag_CL03T00C08_V1
Bacteroides fragilis	CL03T12C07	GCA_000273115.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000273115.1_Bact_frag_CL03T12C07_V1
Bacteroides fragilis	CL05T12C13	GCA_000273135.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000273135.1_Bact_frag_CL05T12C13_V1
Bacteroides fragilis	CL07T12C05	GCA_000273155.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000273155.1_Bact_frag_CL07T12C05_V1
Bacteroides fragilis	CL05T00C42	GCA_000273765.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000273765.1_Bact_frag_CL05T00C42_V1
Bacteroides fragilis	HMW 615	GCA_000297735.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000297735.1_Bact_frag_HMW_615_V1
Bacteroides fragilis	DCMOUH0042B	GCA_000724795.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000724795.1_DCMOUH0042B1.0
Bacteroides fragilis	2-F-2 #7	GCA_000598145.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598145.1_ASM59814v1

Bacteroides fragilis	3976T7	GCA_000598165.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598165.1_ASM59816v1
Bacteroides fragilis	3986T(B)10	GCA_000598185.2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598185.2_ASM59818v2
Bacteroides fragilis	3988 T1	GCA_000598205.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598205.1_ASM59820v1
Bacteroides fragilis	3996 N(B) 6	GCA_000598225.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598225.1_ASM59822v1
Bacteroides fragilis	DS-166	GCA_000598245.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598245.1_ASM59824v1
Bacteroides fragilis	1007-1-F #8	GCA_000598265.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598265.1_ASM59826v1
Bacteroides fragilis	2-F-2 #5	GCA_000598285.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598285.1_ASM59828v1
Bacteroides fragilis	3774 T13	GCA_000598305.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598305.1_ASM59830v1
Bacteroides fragilis	3783N1-2	GCA_000598325.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598325.1_ASM59832v1
Bacteroides fragilis	3783N2-1	GCA_000598345.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598345.1_ASM59834v1
Bacteroides fragilis	3988T(B)14	GCA_000598365.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598365.1_ASM59836v1
Bacteroides fragilis	3998 T(B) 4	GCA_000598385.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598385.1_ASM59838v1
Bacteroides fragilis	34-F-2 #13	GCA_000598425.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598425.1_ASM59842v1
Bacteroides fragilis	3986 N(B)19	GCA_000598445.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598445.1_ASM59844v1
Bacteroides fragilis	3986 T(B)9	GCA_000598465.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598465.1_ASM59846v1
Bacteroides fragilis	3998T(B)3	GCA_000598485.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598485.1_ASM59848v1
Bacteroides fragilis	DS-208	GCA_000598505.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598505.1_ASM59850v1
Bacteroides fragilis	J-143-4	GCA_000598525.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598525.1_ASM59852v1
Bacteroides fragilis	1007-1-F #4	GCA_000598545.2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598545.2_ASM59854v2
Bacteroides fragilis	3397 N2	GCA_000598565.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598565.1_ASM59856v1
Bacteroides fragilis	3725 D9(v)	GCA_000598585.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598585.1_ASM59858v1
Bacteroides fragilis	3783N1-8	GCA_000598605.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598605.1_ASM59860v1
Bacteroides fragilis	B1 (UDC16-1)	GCA_000598625.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598625.1_ASM59862v1
Bacteroides fragilis	J38-1	GCA_000598645.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598645.1_ASM59864v1
Bacteroides fragilis	S23 R14	GCA_000598665.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598665.1_ASM59866v1
Bacteroides fragilis	1007-1-F #10	GCA_000598685.2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598685.2_ASM59868v2
Bacteroides fragilis	1009-4-F #10	GCA_000598705.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598705.1_ASM59870v1
Bacteroides fragilis	3719 T6	GCA_000598725.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598725.1_ASM59872v1
Bacteroides fragilis	S24L26	GCA_000598745.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598745.1_ASM59874v1
Bacteroides fragilis	S38L3	GCA_000598765.2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598765.2_ASM59876v2
Bacteroides fragilis	I1345	GCA_000598785.2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598785.2_ASM59878v2
Bacteroides fragilis	Ds-233	GCA_000598805.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598805.1_ASM59880v1
Bacteroides fragilis	2-F-2 #4	GCA_000598825.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598825.1_ASM59882v1
Bacteroides fragilis	3719 A10	GCA_000598845.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598845.1_ASM59884v1
Bacteroides fragilis	3-F-2 #6	GCA_000598865.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598865.1_ASM59886v1
Bacteroides fragilis	1007-1-F #9	GCA_000598885.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598885.1_ASM59888v1
Bacteroides fragilis	20793-3	GCA_000598905.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598905.1_ASM59890v1
Bacteroides fragilis	3397 N3	GCA_000598925.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598925.1_ASM59892v1
Bacteroides fragilis	3986 N(B)22	GCA_000598945.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598945.1_ASM59894v1
Bacteroides fragilis	3986 T(B)13	GCA_000598965.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598965.1_ASM59896v1
Bacteroides fragilis	A7 (UDC12-2)	GCA_000598985.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000598985.1_ASM59898v1
Bacteroides fragilis	S24L15	GCA_000599005.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000599005.1_ASM59900v1
Bacteroides fragilis	S36L5	GCA_000599025.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000599025.1_ASM59902v1
Bacteroides fragilis	S6R5	GCA_000599045.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000599045.1_ASM59904v1
Bacteroides fragilis	3783N1-6	GCA_000599065.2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000599065.2_ASM59906v2
Bacteroides fragilis	DS-71	GCA_000599085.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000599085.1_ASM59908v1
Bacteroides fragilis	S13 L11	GCA_000599105.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000599105.1_ASM59910v1
Bacteroides fragilis	S36L11	GCA_000599125.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000599125.1_ASM59912v1
Bacteroides fragilis	1007-1-F #7	GCA_000599145.2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000599145.2_ASM59914v2
Bacteroides fragilis	3397 T14	GCA_000599165.2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000599165.2_ASM59916v2
Bacteroides fragilis	3976T8	GCA_000599185.2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000599185.2_ASM59918v2
Bacteroides fragilis	Korea 419	GCA_000599205.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000599205.1_ASM59920v1
Bacteroides fragilis	S6L3	GCA_000599225.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000599225.1_ASM59922v1
Bacteroides fragilis	S6R6	GCA_000599245.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000599245.1_ASM59924v1
Bacteroides fragilis	1007-1-F #3	GCA_000599265.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000599265.1_ASM59926v1
Bacteroides fragilis	1009-4-F #7	GCA_000599285.2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000599285.2_ASM59928v2
Bacteroides fragilis	S23L24	GCA_000599305.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000599305.1_ASM59930v1
Bacteroides fragilis	S24L34	GCA_000599325.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000599325.1_ASM59932v1
Bacteroides fragilis	S36L12	GCA_000599345.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000599345.1_ASM59934v1
Bacteroides fragilis	S38L5	GCA_000599365.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000599365.1_ASM59936v1
Bacteroides fragilis	S6L8	GCA_000599385.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000599385.1_ASM59938v1
Bacteroides fragilis	1007-1-F #5	GCA_000601035.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000601035.1_ASM60103v1
Bacteroides fragilis	S23L17	GCA_000601055.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000601055.1_ASM60105v1
Bacteroides fragilis	S6R8	GCA_000601075.2	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000601075.2_ASM60107v2
Bacteroides fragilis	1007-1-F #6	GCA_000601095.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000601095.1_ASM60109v1
Bacteroides fragilis	3986 N3	GCA_000601115.1	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000601115.1_ASM60111v1