Acquired interbacterial defense systems protect against interspecies antagonism in the human gut microbiome

Benjamin D. Ross[1,*], Adrian J. Verster[2,*], Matthew C. Radey[1], Danica T. Schmidtke[1], Christopher E. Pope[3], Lucas R. Hoffman[1,3,4], Adeline Hajjar[5], S. Brook Peterson[1], Elhanan Borenstein[2,6,7,8,9,†] and Joseph D. Mougous[1,10,11,†]

[1]Department of Microbiology, University of Washington, Seattle, WA 98195, USA; [2]Department of Genome Sciences, University of Washington, Seattle, WA, 98195, USA; [3]Department of Pediatrics, University of Washington, Seattle, WA, 98195, USA; [4]Seattle Children's Hospital, Seattle, WA, 98105; [5]Department of Comparative Medicine, University of Washington, Seattle, WA 98195, USA; [6]Department of Computer Science and Engineering, University of Washington, Seattle, WA 98195, USA; [7]Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv 6997801, Israel; [8]Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 6997801, Israel; [9]Santa Fe Institute, Santa Fe, NM 87501, USA; [10]Department of Biochemistry, University of Washington, Seattle, WA, 98195, USA; [11]Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA

*Equal contribution. [†]Correspondence: mougous@uw.edu, elbo@uw.edu

Abstract

The genomes of bacteria derived from the gut microbiota are replete with pathways that mediate contact-dependent interbacterial antagonism. However, the role of direct interactions between co-resident microbes in driving microbiome composition is not well understood. Here we report the widespread occurrence of acquired interbacterial defense (AID) gene clusters in the human gut microbiome. These clusters are found on predicted mobile elements and encode arrays of immunity genes that confer protection against interbacterial toxin-mediated antagonism *in vitro* and in gnotobiotic mice. We find that *Bacteroides ovatus* strains containing AID systems that inactivate *B. fragilis* toxins delivered between cells by the type VI secretion system are enriched in samples lacking detectable *B. fragilis*. Moreover, these strains display significantly higher abundance in gut metagenomes than strains without AID systems. Finally, we identify a recombinase-associated AID subtype present broadly in Bacteroidales genomes with features suggestive of active gene acquisition. Our data suggest that neutralization of contact-dependent interbacterial antagonism via AID systems plays an important role in shaping human gut microbiome ecology.

Polymicrobial environments contain a plethora of biotic and abiotic threats to their inhabitants. Bacterial survival in these settings necessitates elaborate defensive mechanisms. Some of these are basal and protect against a wide range of threats, whereas others, for instance CRISPR-Cas, represent adaptations unique to the specific threats encountered by a bacterial lineage [1-3]. The density of bacteria in the mammalian gut microbiome can exceed 10^{11} gm^{-1}; therefore, overcoming contact-dependent interbacterial antagonism is likely a major hurdle to survival in this ecosystem [4]. The type VI secretion system (T6SS) is a pathway predicted to be widely utilized by gut bacteria to mediate the delivery of toxic effector proteins to neighboring cells [5-8]. While kin cells are innately resistant to these effectors via cognate immunity proteins, it is unknown whether non-self cells in the gut can escape intoxication.

To identify potential mechanisms of defense against T6S-delivered interbacterial effectors, we mined a large collection of shotgun metagenomic samples from the human gut microbiome for sequences homologous to known immunity genes [9-11]. We first focused our efforts on *Bacteroides fragilis*, a bacterium belonging to the order Bacteroidales, which harbors a well described and diverse repertoire of effector and cognate immunity genes (Supplemental Table 1) [5, 6, 12]. As expected for genes predicted to reside within the *B. fragilis* genome, sequences mapping to these effector and immunity loci were detected at an abundance similar to that of *B. fragilis* species-specific marker genes in many microbiome samples (Fig. 1A-B). However, in a second subset of samples, immunity genes were detected at an abundance significantly higher than expected given the abundance of *B. fragilis*, consistent with a prior observation that immunity genes can be encoded outside of T6SS loci (Fig. 1B) [6]. Finally, we identified a third subset of samples wherein immunity gene sequences were detected in the absence of *B. fragilis* or cognate effector gene sequences. These latter sequences included close homologs of 11 of 14 unique immunity genes encoded by *B. fragilis* (Fig. 1C).

The detection of *B. fragilis* immunity gene homologs in samples in which we were unable to detect *B. fragilis* strongly suggests that these elements are encoded by other bacteria in the gut. To identify these bacteria, we assembled full-length predicted immunity genes from the metagenomic sequencing reads of individual microbiomes. We limited this assembly to homologs of immunity genes 6 and 7 (i6 and i7) – the most prevalent immunity genes detected in samples lacking *B. fragilis*. Clustering of the recovered homologs showed that the majority of sequences distribute into a limited number of discrete clades that differ by one or more nucleotide substitutions; three and four clades were observed for immunity genes 6 and 7, respectively (i6:cI-cIII and i7:cI-cIV) (Fig. 2A-B and Supplemental Table 2). A comparison of these immunity sequences to publicly available bacterial genomes revealed two clades matching cognate immunity genes in *B. fragilis* genomes (i6:cI and i7:cI).

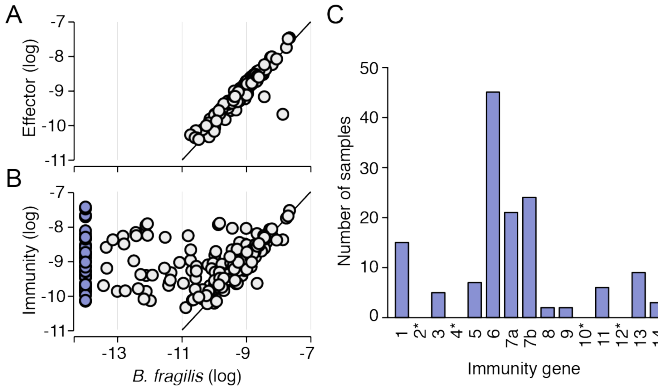


Fig. 1: T6SS orphan immunity genes are widespread and abundant in human microbiomes. A-B. Normalized abundance of *B. fragilis* T6SS effector (**A**) or immunity genes (**B**) plotted against the normalized abundance of *B. fragilis* marker genes in adult microbiome samples ($n = 553$). Abundances are log10 scaled with a pseudo count of 10^{-14}. Blue points denote samples with undetectable *B. fragilis*. **C.** The number of adult human gut microbiome samples with undetectable *B. fragilis* (blue points in B) in which the indicated immunity genes (1-14, GA3_i1-14 from ref [5]) can be detected at an 80% nucleotide identity threshold. Asterisks indicate gene is undetected in samples.

Additionally, we found sequences homologous to several clades in other *Bacteroides* species, including *B. ovatus* (i6:cIII and i7:cIII) and *B. xylanisolvens* (i6:cII), which we previously found do not contain the cognate T6SS effector genes [5]. To further define the species encoding these sequences, we used a simplified linear model to identify *Bacteroides* spp whose abundance in microbiome samples best fits that of each immunity sequence clade (Supplemental Fig. 1A-B). As predicted by our genomic-based analysis, immunity clades i6:cIII and i7:cIII are best explained in gut metagenomes by *B. ovatus*, and i6:cII by *B. xylanisolvens* (Fig. 2C-D, F). This analysis further suggested that *B. thetaiotaomicron* is likely to harbor sequences from i7:cII (Fig. 2E). Based on these observations, we conclude that T6S immunity genes derived from *B. fragilis* are encoded by other species of *Bacteroides* and are widespread in human gut microbiomes. We hypothesize that these genes serve an adaptive role in the gut by providing defense against *B. fragilis* interspecies antagonism.

We also identified orphan immunity genes in *B. fragilis* genomes identical to a sequence clade that diverges from cognate immunity (i7:cIV) (Fig. 2B) [6]. Abundance correlation analyses supported the presence of this sequence clade in *B. fragilis* strains in microbiome samples (Fig. 2G). Orphan immunity genes corresponding to effectors harbored by strains of the same species have also been observed in other bacteria [13-17] . These genes are thought to arise through the degeneration of T6S effector–immunity loci, recombination, or gene duplication. In contrast, orphan immunity genes in other species, such as those we identified in *B. ovatus* and *B. xylanisolvens*, were likely acquired laterally from *B. fragilis*.

To gain insight into the function of orphan immunity genes, we examined their genomic context. This revealed that orphan i6 and i7 genes belonging to clades i6:cIII and i7:cIII-IV are located together within a discrete gene cluster (Fig. 2H). The organization and nucleotide sequence of this cluster, which we termed the AID-1 (acquired interbacterial defense 1) system is conserved between *Bacteroides* spp, and we found that it resides on a predicted mobile integrative and conjugative element (ICE) (Supplemental Fig. 1C) [18]. Interestingly, AID-1 systems also contain distant homologs and pseudogenized remnants of additional *B. fragilis* T6S immunity genes, including i4, i5, i9, i11, and i14 (Fig. 2H, I, Supplemental Table 3). Owing to their sequence divergence from cognate immunity genes, these loci were not detected in our initial, high-stringency search for and quantification of immunity gene sequence reads in gut metagenomes. However, query of a database of non-redundant complete gene sequences compiled from human gut microbiome samples showed that distant homologs of numerous *B. fragilis* immunity genes occur in the human gut microbiome (Supplemental Fig. 2A) [19]. Assembly of sequence scaffolds from metagenomic data confirmed the presence of AID-1 gene clusters in multiple gut microbiome samples (Supplemental Fig. 2B). In *B. xylanisolvens*, we found that genes belonging to i6:cII are located in a unique, but analogous gene context we designate as the AID-2 system; they reside on an apparent transposable element adjacent to a homolog of i5 (Fig. 2H) [20]. Together, these findings suggest that mobile orphan immunity islands provide *Bacteroides* species defense against *B. fragilis* T6S.

We next sought to define the phenotypic implications of orphan immunity genes of *Bacteroides* spp during competition with *B. fragilis*. *B. fragilis* 9343 encodes the cognate effectors for i6 and i7, and prior data demonstrate that the corresponding toxins, e6 and e7, respectively, efficiently antagonize assorted *Bacteroides* spp *in vitro* and in gnotobiotic mice [6]. We thus employed this strain in growth competition assays against

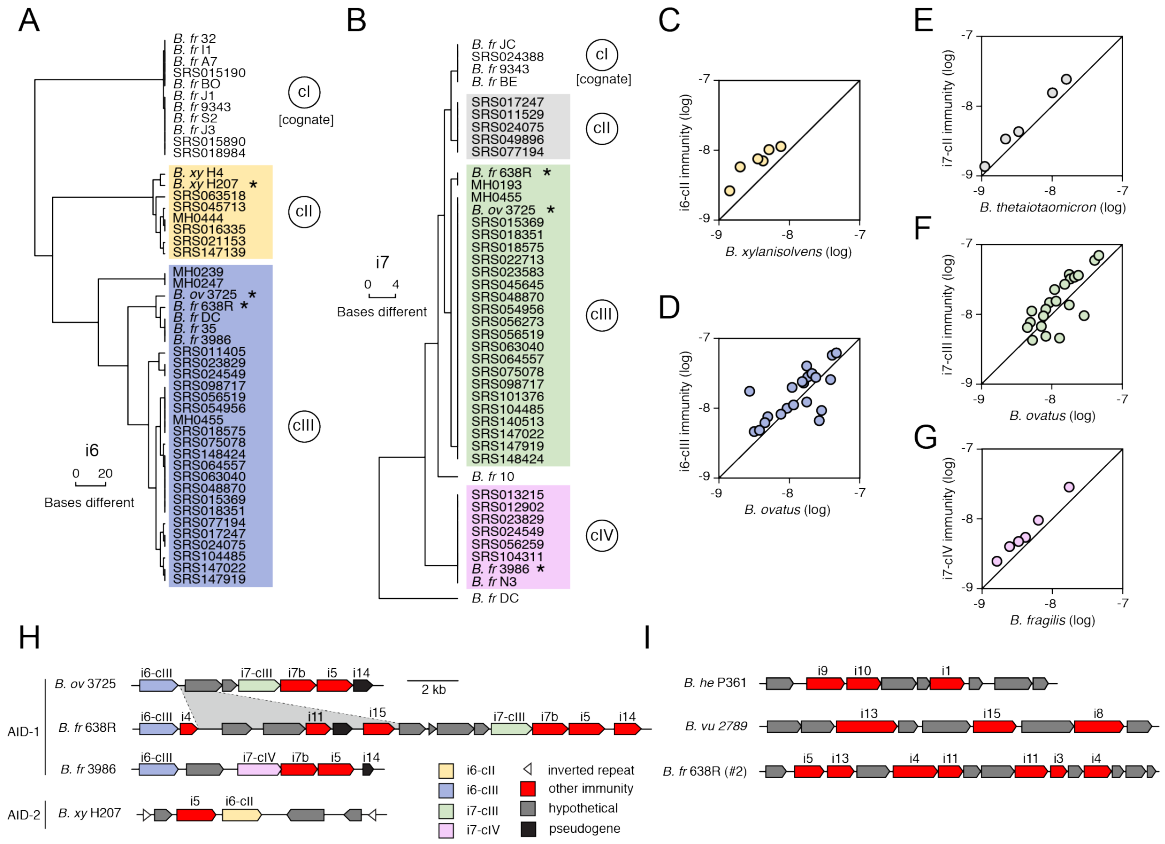


Fig. 2: T6SS orphan immunity is encoded by multiple species in the human microbiome. A-B. Dendrogram depicting hierarchical clustering of i6 (**A**) or i7a (**B**) gene sequences extracted from genomes (n = 15, i6; n =9, i7) and metagenomes (n = 32, i6; n = 34, i7). For genomes, immunity genes were identified using BLAST, while for metagenomes gene sequences were constructed from a pileup of shotgun reads aligned to the cognate reference sequence. Only samples with at least 10X coverage over 90% of the length of the gene were included. Colored boxes; clusters of identical sequences representing clades designated by circled labels (i6, cI-III; i7, cI-IV). Asterisks; strains for which orphan immunity gene clusters are shown in H. Strain name abbreviations defined in Supplemental Data Table 2. **C-G.** Normalized abundance in adult gut microbiome samples of genes from the indicated orphan immunity clades (colors as in A and B) plotted against normalized abundances of the marker genes of the indicated species. Species shown are those with abundance most closely matching that of the indicated immunity genes (see Supplemental Fig. 1A). Gene abundance calculated as in Fig. 1. **H-I**. To-scale depiction of representative gene clusters containing orphan immunity genes from the indicated *Bacteroides* spp (red, non i6 and i7 immunity genes; grey, genes encoding hypothetical proteins; black, pseudogenes; other colors, clades shown A-B). Region of difference between otherwise homologous clusters in *B. ovatus* 3725 and *B. fragilis* 638R is highlighted.

Bacteroides spp bearing orphan immunity genes or derivative strains containing deletions thereof. These experiments showed that in both *B. ovatus* and *B. fragilis*, AID-1 system genes belonging to i6:cIII and i7:cIII grant immunity against corresponding T6S effectors (Fig. 3A, Supplemental Fig. 3A). We similarly found that an i6:cII gene from an AID-2 system in *B. xylanisolvens* affords this bacterium protection against e6 of *B. fragilis* 9343 (Fig. 3B). In total, these data show that the orphan immunity genes of multiple *Bacteroides* spp – localized to clusters within predicted ICE or transposable elements – can confer protection against effectors delivered by the T6SS of *B. fragilis*.

B. fragilis is typically found as a clonal population in the human gut microbiome, and recent studies suggest that this is in part due to active strain exclusion via the T6SS [5, 21, 22]. However, in gnotobiotic mouse colonization experiments, certain *B. fragilis* strain pairs inexplicably co-exist [7]. We noted that one such pair corresponds to *B. fragilis* 9343 and *B. fragilis* 638R, the latter of which contains an AID-1 which includes i6:cIII and i7:cIII. To determine whether our *in vitro* results with these strains extend to a more physiological setting, we measured the fitness contribution of the orphan immunity genes encoded by *B. fragilis* 638R following pre-colonization of the murine gut with *B. fragilis* 9343 (Fig. 3C, Supplemental Fig. 3B, C). Our results indicated that the cumulative protection afforded by orphan i6 and i7 genes underlies the capacity of *B. fragilis* 638R to persist during T6S-mediated antagonism *in vivo*.

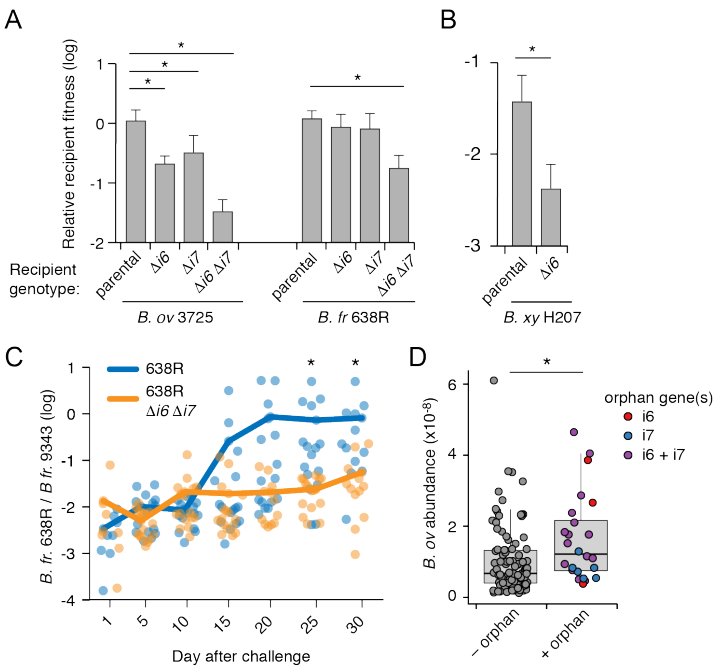


Fig. 3: Orphan immunity genes enhance the fitness of *Bacteroides* strains *in vitro* and *in vivo*. A-B. T6SS-dependent competitiveness of parental strains of *B. ovatus* 3725 D1 iv and *B. fragilis* 638R (**A**) or *B. xylanisolvens* H207 (**B**) harboring AID-1 or AID-2 or those bearing in-frame deletions of indicated orphan immunity genes during *in vitro* growth competition experiments with *B. fragilis* NCTC 9343. Competitive index was determined by comparison of final and initial c.f.u. ratios of indicated strains (recipient to donor ratio), normalized to the corresponding competitions with *B. fragilis* NCTC 9343 lacking *tssC* (T6-inactive). Data represent mean ± s.d. for three technical replicates from each of at least three biological replicates. Asterisks in A-B indicate statistically significant differences between indicated mean values (t-test, $p < 0.01$). **C.** Outcome of pairwise competition between the indicated *B. fragilis* strains in germ-free mice (blue, parental NCTC 9343 vs. parental 638R; orange, parental NCTC 9343 vs. 638R lacking orphan immunity). Mice were pre-colonized with NCTC 9343 for one week prior to challenge with 638R; points depict the 9343/638R ratio as determined by selective plating following collection of fresh fecal samples on days indicated. Lines represent means for each group. Asterisk indicates statistically significant difference (Wilcoxon rank sum test for each timepoint, * $P < 0.01$; N=8 mice per group from two independent experiments, data pooled from consecutive timepoints). **D.** Abundance of *B. ovatus* in samples lacking detected orphan immunity genes (-) and samples in which the indicated orphan immunity genes were assigned to *B. ovatus* (+; colors as in Fig. 2). Asterisks indicate statistical significance (Wilcoxon rank sum test, * $P<0.01$, n =125 for i6 and n = 149 for i7).

Deciphering the contribution of individual gene products, or even whole pathways, to bacterial fitness in complex microbial communities is challenging; however, consideration of such parameters is critical for efforts to engineer stable functional consortia. We reasoned that the ability to identify orphan immunity genes in human gut metagenomes, coupled with our capacity to infer their organismal source, provided an opportunity to measure the impact of these defensive factors on competitiveness in the gut. To this end, we compared the abundance of *B. ovatus* strains with and without i6 and i7 orphan immunity genes in human gut metagenome samples. Remarkably, we found that the average abundance of *B. ovatus* strains with orphan immunity genes significantly exceeds that of those without (Fig. 3D). This finding suggests that the acquisition of i6 and i7 allows *B. ovatus* to increase its niche, likely via the displacement of *B. fragilis*. This apparent selective advantage conferred by orphan immunity genes offers an explanation for the widespread occurrence of these factors on transmissible elements.

Given the benefit that acquisition of immunity genes against *B. fragilis* effectors confers to *B. ovatus*, we hypothesized that this mechanism of inhibiting interbacterial antagonism should extend to effectors produced by other species. We previously reported evidence that *B. fragilis* is antagonized by other *Bacteroides* spp in the human gut microbiome [5]. In addition to the T6SS present exclusively in *B. fragilis*, this species and other *Bacteroides* spp can possess a second T6SS with a distinct and non-overlapping repertoire of effector and immunity genes [5, 12]. Therefore, we searched *B. fragilis* genomes for sequences homologous to this group of immunity genes. In 29 of the 122 available *B. fragilis* genomes, we identified apparent orphan homologs of these immunity genes grouped within related gene clusters (Fig. 4A). While analogous to the AID-1 and -2 systems, these clusters have several unique characteristics including conservation of a gene encoding a predicted XerD-family tyrosine recombinase and repetitive intergenic sequences reminiscent of those present in integrons (Supplemental Fig. 4A) [23, 24]. These often-large gene clusters, hereafter referred to as recombinase-associated AID (*r*AID-1) systems, can exceed 16 kb and contain up to 31 genes with varying degrees of homology to T6S immunity genes and predicted immunity genes associated with other interbacterial antagonism pathways (Supplemental Fig. 4B, C, Supplemental Table 4) [15].

Using the shared characteristics of *B. fragilis* *r*AID-1 systems, we searched for related gene clusters across sequenced Bacteroidales genomes. We found that over half of sequenced bacteria belonging to this order possess a *r*AID-1 system. In sum, these gene clusters contain 579 unique genes, encompassing homologs of 25 Bacteroidales T6S immunity genes (Fig. 4A and Supplemental Table 5). Furthermore, characterization of an orphan immunity gene from the *B. fragilis* 9343 *r*AID-1 cluster indicates that these genes can neutralize heterologously-expressed effectors and those delivered during T6-mediated interbacterial antagonism (Fig. 4B, C). Taken together with our functional characterization of AID-1 and AID-2, these findings suggest that acquisition and maintenance of consolidated orphan immunity determinants is a ubiquitous mechanism for promoting bacterial competitiveness in the human gut microbiome.

Mounting evidence suggests that competitive interactions between bacteria predominate in many environments [25]. This

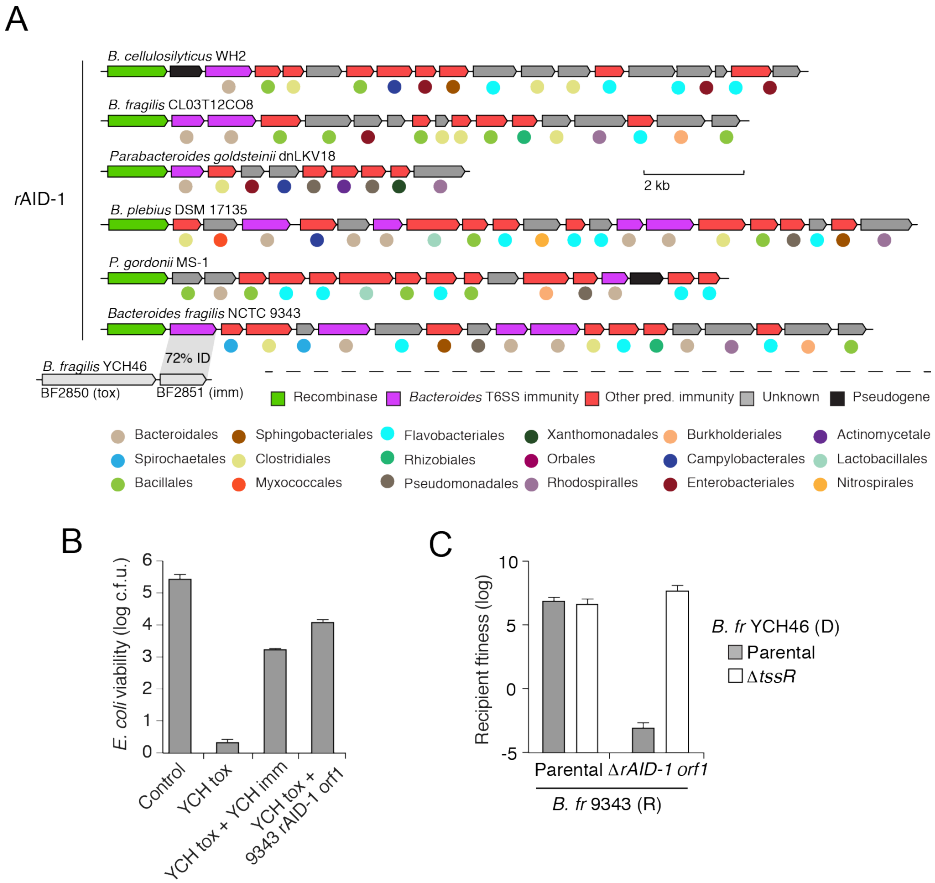


Fig. 4: Functional interbacterial immunity harbored within recombinase-associated acquired interbacterial defense system (*r*AID-1). A. To-scale depiction of *r*AID-1 clusters from *B. fragilis* NCTC 9343. Cluster genes were assigned to functional immunity classes (indicated by gene coloring) via profile HMM scans and BLAST analysis using a curated database of Bacteroidales T6SS immunity genes [5, 15]. Taxonomic association of the top non-cluster homolog identified via tblastn against the nr database (colored circle below each gene, taxonomic order). Homology (% amino acid identity) between *orf1* of the *B. fragilis* NCTC 9343 *r*AID-1 cluster and a T6SS cognate immunity gene from *B. fragilis* YCH46 is indicated by gray shading. **B.** Viable *E. coli* cells recovered from cultures carrying plasmids expressing the indicated proteins (c.f.u., colony-forming units). **C.** Relative competitiveness of the indicated donor (*B. fragilis* YCH46) and recipient (*B. fragilis* NCTC 9343) strains during *in vitro* growth competition experiments. Competitive index was determined by comparison of final and initial c.f.u. ratios of indicated strains (recipient to donor ratio). Data represent mean ± s.d. for three technical replicates from two biological replicates. Asterisks indicate statistically significant differences between indicated mean values (t-test, $p < 0.01$).

evolutionary pressure has undoubtedly led to the wide dissemination of idiosyncratically orphaned immunity genes predicted to provide resistance to diverse antagonistic pathways [14, 15, 17, 26, 27]. Modeling studies predict that interbacterial antagonism is a critical contributor to the maintenance of a stable gut community [28]. Our findings reveal that a corollary of the pervasiveness of antagonistic mechanisms is strong selective pressure for genes that can provide protection against attack, establishing a molecular arms race that has led to diversification and expansion of T6S effectors. Deciphering the linkage between orphan immunity genes and the bacteria harboring the cognate effectors has the promise of providing a window into the physical connectivity of bacteria in the gut microbiome.

It is now appreciated that phage defense mechanisms, including the adaptive system, CRISPR-Cas, are critical for bacteria to cope with the omnipresent threat and deleterious outcome of phage infection [3]. However, the ubiquity of interbacterial antagonistic systems suggests that in most habitats, bacteria are equally, or perhaps more likely to be subject to attack and potential cell death via the action of other bacteria [15, 16, 29, 30]. Despite this, consolidated interbacterial defense systems have not been investigated experimentally, nor their ecological consequences understood. Our characterization of elaborate, widespread AID systems encoded by members of the human gut microbiota appears to reconcile these observations and demonstrate that the neutralization of contact-dependent interbacterial antagonism can be a critical mechanism for survival in polymicrobial environments. Additionally, it suggests that analogous to the immune system of vertebrates, that of bacteria includes arms specialized in viral or bacterial defense.

Methods

Microbiome Data

Metagenomic data from healthy adults were obtained from a number of large-scale sequencing projects. We specifically obtained 147 samples from the Human Microbiome Project 1.0, 100 samples from HMP 1.2, 99 and 207 samples from two different MetaHIT datasets [9-11, 32]. In addition, we obtained a database of genes identified from 1267 assembled metagenomes as part of the Integrated Gene Catalogue (IGC) [19].

Analysis of gene and species abundances in microbiome samples

We compiled a list of T6SS immunity and effector genes [5]. We additionally compiled a list of species-specific marker genes for all *Bacteroides* species obtained from MetaPhlAn 2.0 [33]. In order to determine the abundance of a given immunity,

effector, or marker gene in each metagenomic sample, single end metagenomic reads were aligned to gene sequences using bowtie2, allowing for one mismatch in the seed [34]. We counted the number of reads that aligned to each such gene with at least 80% nucleotide identity (to encompass divergent orphan immunity gene sequences) and minimum mapping quality of 20. The abundance of a gene was calculated as the number of reads aligned to this gene, normalized by the gene length and by the library size. For each species, the average gene level abundance of all species-specific marker genes was used to assess the species abundance.

Orphan immunity phylogenetic analysis

Filtered reads derived from human shotgun microbiome datasets were aligned using bowtie2 as described above and subsequently converted to a pileup using samtools with parameters --excl-flags UNMAP,QCFAIL,DUP -A -q0 -C0 -B [34, 35]. A sequence corresponding to the most abundant version of the immunity gene in the sample was reconstructed from that pileup as follows. First, 50 bases from the start and end were trimmed due to a propensity for low coverage. Second, at all sites with at least 10X coverage the base was set to the major allele. Sites with <10X coverage were assigned an ambiguous base. Finally, we only kept the reconstructed sequence in metagenomic samples where at least 90% of the sequence had >10X coverage. The number of SNPs between all immunity sequences, both from metagenomic samples and from *Bacteroides* genomes, was calculated and used to populate a distance matrix. Since obtained distances were small (e.g., a single base difference), we used hierarchical clustering (with complete linkage), rather than standard phylogenetic reconstruction methods, to visualize the relatedness between different sequences.

Assigning orphan immunity sequences to bacterial species

We aimed to identify the species most likely to encode the immunity gene in each cluster of identical sequences reconstructed from metagenomes. Only clusters with at least 3 sequences were used, to ensure statistical confidence. The abundance of each species was assessed based on species-specific marker genes as described above. We specifically employed a simple linear model that assumed that only a single species encodes the immunity gene. We further assumed a one to one relationship between species marker gene abundance and orphan immunity gene abundance, and accordingly fixed the intercept at zero and allowed a single species with a slope of one. The fit of the model for each species was calculated as the mean square error over all samples. The most likely species to encode the immunity gene was determined by the minimum mean squared error.

Assembly of orphan immunity sequences from metagenomes

Paired-end metagenomic sequencing data was assembled using SoapDeNovo2 with a kmer length of 63 and an average insert size of 200 [36]. BLAST was used to identify the contig that contained the orphan immunity gene, and GeneMarkS was used predict protein coding genes [37].

Bacterial culture conditions

Anaerobic culturing procedures were performed either in an anaerobic chamber (Coy Laboratory Products) filled with 70% N_2, 20% CO_2 and 10% H_2, or in Becton Dickson BBL EZ GasPak chambers. *E. coli* EC100D λ pir and S17-1 λ pir strains were grown aerobically at 37° C on lysogeny broth agar. Unless otherwise noted, *Bacteroides* strains were cultured under anaerobic conditions on brain heart infusion agar (BHI; Becton Dickinson) supplemented with 50 μg/mL hemin and 10% sodium bicarbonate [38]. Antibiotics and chemicals were added to media as needed at the following concentrations: carbenicillin 150 μg/mL, gentamicin 60 μg/mL, erythromycin 12.5 μg/mL, tetracycline 6 μg/mL, floxuridine (FUdR) 200 μg/mL.

Genetic techniques

Standard molecular procedures were employed for creation, maintenance and *E. coli* transformation of plasmids. All primers used in this study were synthesized by Integrated DNA Technologies (IDT). Phusion polymerase, restriction enzymes, T4 DNA ligase, and Gibson Assembly Reagent were obtained from New England Biolabs (NEB). A comprehensive list of primers, plasmids, and strains are provided (Supplemental Data Table 6). Deletion of the gene encoding thymidine kinase in *B. fragilis*, *B. ovatus*, and *B. xylanisolvens* strains was performed by cloning respective genomic flanking regions into the vector pKNOCK as previously described [39]. Briefly, pKNOCK-*tdk* plasmids were mobilized into *Bacteroides* strains via overnight aerobic mating with *E. coli*. Integrants were isolated by plating on selective media, were passaged once without antibiotics to allow for plasmid recombination, and plated for counter selection on FUdR. Recovered single colonies were patched onto selective media to ensure loss of pKNOCK, and disruption of *tdk* confirmed by PCR . Subsequent deletion of orphan immunity genes was performed in □*tdk* strains via a similar counter selection strategy, except employing the suicide plasmid pExchange in place of pKNOCK [6]. Genomic deletions were confirmed by PCR. Gene complementation was performed by cloning genes into pNBU2-erm_us1311 for constitutive expression [40].

Isolation of *Bacteroides* strains from fecal samples

Fecal samples from healthy infants used for strain isolation were collected as part of a prior study approved by the Seattle Children's Hospital Institutional Review Board [41, 42]. Frozen stool samples stored at -80° C were manually homogenized, serially diluted in tryptone yeast glucose (TYG) broth, and plated under anaerobic conditions on *Bacteroides* bile esculin (BBE) agar plates (Oxyrase, Inc). Colonies which exhibited esculin hydrolysis as indicated by the production of black pigment on BBE agar were sub-cultured in TYG broth with the addition of 60ug/mL gentamicin until stationary phase and then

were frozen at -80° C following the addition of sterile glycerol to 20% final concentration. Single colonies restreaked from these stocks were subsequently screened by PCR with primers targeting the orphan i6 gene as assembled from metagenomic short read sequence data [42].

Genome sequencing

Genomic DNA used for Illumina sequencing was prepared by harvesting *Bacteroides* strains grown overnight on BHIS blood agar plates. Cells resuspended from plates were washed in phosphate buffered saline before DNA extraction with the Qiagen DNeasy Blood and Tissue Kit. Sequencing was performed on an Illumina MiSeq using the V3 Reagent kit at the Northwest Genomics Center sequencing facility at the University of Washington. Assembly was done with paired end Illumina reads using SPAdes version 3.7.1 in 'careful' mode [43]. We used a 93 K-mer setting for *B. xylanisolvens* H207 and 99 K-mer setting for *B. ovatus* 3725 D1 iv. Open reading frames on genome scaffolds were annotated using Prokka 1.12 [44]. AID clusters often appear in highly repetitive genomic contexts (e.g. mobile elements) and are often split into multiple scaffolds in reference genomes. To compensate for this, we additionally performed long read sequencing via PacBio on a subset of genomes. To this end, high-molecular weight DNA was extracted using the Qiagen Genomic-tip Kit and sequenced by SNPsaurus (Eugene, OR) using a PacBio Sequel. PacBio assemblies were generated using Canu 1.7 [45]. Species identification was performed by blast searches with species-specific marker genes [33]. Whole genome sequencing data generated in the course of this study has been deposited at the Sequence Read Archive under BioProject Accession PRJNA484981.

Interbacterial competition assays

Bacteroidales strains were grown on BHIS blood agar plates overnight at 37° C. Bacteria were resuspended from plates in BHIS broth and the optical density of each strain was adjusted to a 10:1 *B. fragilis* NCTC 9343 to competitor ratio (OD_{600} 6.0 to 0.6) for competitions involving *B. xylanisolvens* and *B. ovatus*, or 1:1 ratio for competitions involving *B. fragilis* 638R (OD_{600} 6.0). Equal volumes of each strain at the adjusted OD were mixed and 5ul of bacterial mixtures were spotted onto pre-dried BHIS blood agar plates, in triplicate spots. Competitions were allowed to proceed for 20-24 hours at 37° C under anaerobic conditions before spots were harvested into BHIS broth. Competition outcomes were quantified in one of two ways: 1) by serial dilution for enumeration of colony forming units after plating on BHIS selective plates containing either erythromycin or tetracycline, or 2) purification of total genomic DNA using the Qiagen DNeasy Blood and Tissue Kit and subsequent quantification by qPCR using strain-specific primers (see Supplemental Data Table 6). For antibiotic selection, *B. fragilis* NCTC 9343 was marked with erythromycin resistance by integration of pNBU2-erm at the att1 site [40]. Other strains were either naturally tetracycline resistant, or were marked by integration of pNBU2-tet-BCO1. Strains with insertions of pNBU2 were selected for matching integration by PCR with primers flanking att sites [46].

Gnotobiotic animal studies

Germ-free 6-12 week-old female Swiss Webster mice from multiple litters were randomized, housed simultaneously in pairs in single Techniplast cages, and fed a standard lab diet (Laboratory Autoclavable Rodent Diet 5010, LabDiet), in accordance with guidelines approved by the University of Washington Institutional Animal Care and Use Committee. Blinding was not performed. *Bacteroides fragilis* strains were introduced into mice via oral gavage of 10^8 colony forming units (c.f.u.) suspended in 0.2mL of BHIS broth with 20% glycerol. Challenge with *B. fragilis* 638R strains occurred 7 days following pre-colonization with *B. fragilis* NCTC 9343 strains. Colonization levels by each strain in each mouse were tracked by collection of fecal pellets over a period of 4 weeks, plating on selective BHIS agar plates, and quantification of c.f.u. Differences in the ratio (BF638R/BF9343) of c.f.u. between groups at each timepoint was assessed using Wilcoxon rank sum tests. Non-parametric tests were used following Shapiro-Wilk analysis for normality. Data from consecutive timepoints were merged to increase sample size. Mice were confirmed to be sterile prior to colonization by qPCR with primers targeting the 16S rRNA gene, and free of non-*Bacteroides* contamination by plating fecal pellets on non-selective LB and BHI plates incubated under either anaerobic and aerobic conditions [47].

Assessing fitness contribution of orphan immunity genes in microbiomes

To test for a selective effect of orphan immunity in human microbiomes, we compared metagenomic samples where the orphan immunity gene (I7a or I6) is present and could be assigned to *B. ovatus* (based on our clustergram analysis) to samples containing *B. ovatus* but lacking the orphan immunity gene. To identify such samples in a way that does not bias the abundance of *B. ovatus*, for each immunity gene tested, we first selected the marker gene closest in length to the orphan immunity gene (though, using a different marker gene did not affect the results of this analysis). We then selected a set of samples where this marker gene had at least 10X coverage along 90% of the gene (equivalent to the inclusion criterion for samples with immunity genes above). We next divided the set of samples into those where the orphan immunity gene is inferred to be encoded by *B. ovatus* (based on the clustergram), and those that lack an orphan immunity gene. We finally compared the abundance of *B. ovatus* in these two groups of samples using a Wilcoxon rank sum test.

Bioinformatic analysis of *r*AID clusters

The amino acid sequence of the *B. fragilis* NCTC 9343 polyimmunity-associated XerD-like tyrosine recombinase (BF9343_RS08045) was used as a query against a custom

database of 423 *Bacteroidales* genomes downloaded from GenBank. *r*AID clusters in *Bacteroidales* genomes were identified based upon the following criteria: i) presence of a 5' XerD-like tyrosine recombinase gene encoding a protein with amino acid identity exceeding 44% (corresponding to an e-value of 10^{-100}), ii) 2 or more co-directionally oriented downstream genes which possessed iii) a GC content of 41% or lower. The end of the gene cluster was defined as the stop codon of the last co-directionally oriented gene in the cluster with similar GC content. To identify homologs of genes within *r*AID clusters, open reading frames within the clusters were translated and used as tBlastn queries against the NCBI non-redundant nucleotide database. Top hits from these searches were often genes in other *r*AID clusters; therefore, these hits were discarded. The top non-*r*AID hit from tBlastn searches with an e-value threshold of 10^{-30} was selected as the closest homolog. *r*AID cluster genes were assigned to interbacterial immunity gene families via hmm scans with profiles from Zhang et al [15] with an e-value cutoff of 10^{-3}. *r*AID cluster genes were additionally compared via tblastn with forty-six Bacteroidales T6SS immunity genes from subtypes 1-3 [5, 12] with an e-value cutoff of 10^{-10}. Percent amino acid identity with homologs was assessed if sequences were alignable across ≥80% of their length. Motif enrichment analysis was performed on non-coding sequences within a subset of *r*AID-1 clusters (14 sequences immediately 3' of the recombinase stop codon, and 86 intergenic sequences between *r*AID-1 ORFs), using MEME Suite 5.0.2 and default settings [31].

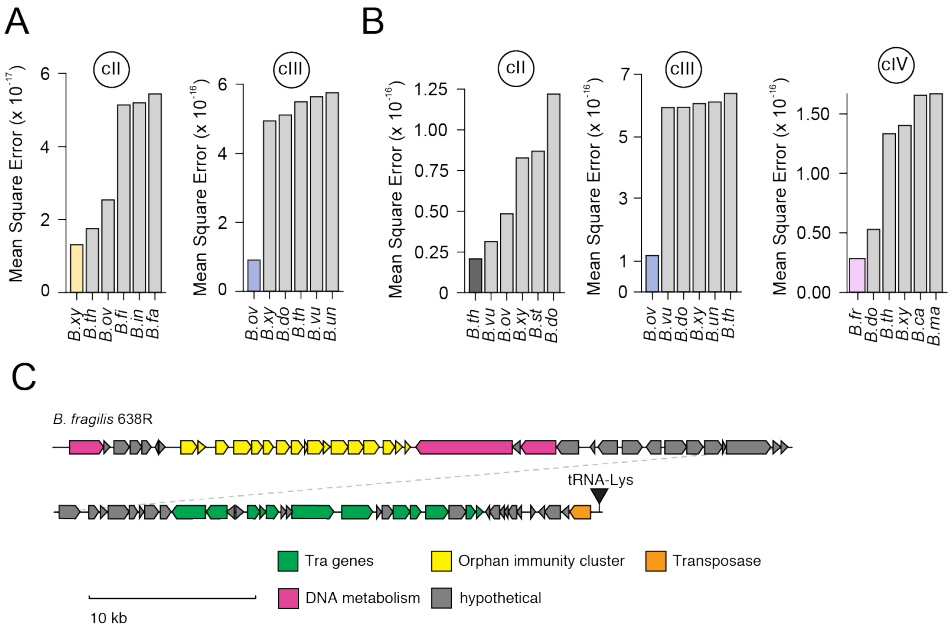
Heterologous expression of *Bacteroides* toxin and immunity genes

To assess the ability of cognate immunity (*B. fragilis* YCH46 BF2851), or orphan immunity (orf1 from the *B. fragilis* 9343 *r*AID-1 cluster, BF9343_1657) to neutralize the toxicity of a *Bacteroides* T6SS toxin (*B. fragilis* YCH46 BF2850) were cloned into expression *E. coli* vectors, with the addition of the *P. aeruginosa* ribosome binding site from *hcp1* [48]. Cultures of *E. coli* 10-beta cells (New England BioLabs) harboring the toxin gene in pSchraB2-V and/or immunity genes in pPSV39 were grown at 37°C in LB containing 50ug/mL trimethoprim and/or 15ug/mL gentamycin in the presence of 10% glucose (to repress expression from pSchraB2-V) to log phase [48, 49]. Cells were then washed, resuspended in LB lacking glucose, 10-fold serially diluted, then plated onto LB agar plates containing 0.05% rhamnose and/or 1mM IPTG. Agar plates were incubated overnight at 37°C before quantification of c.f.u.

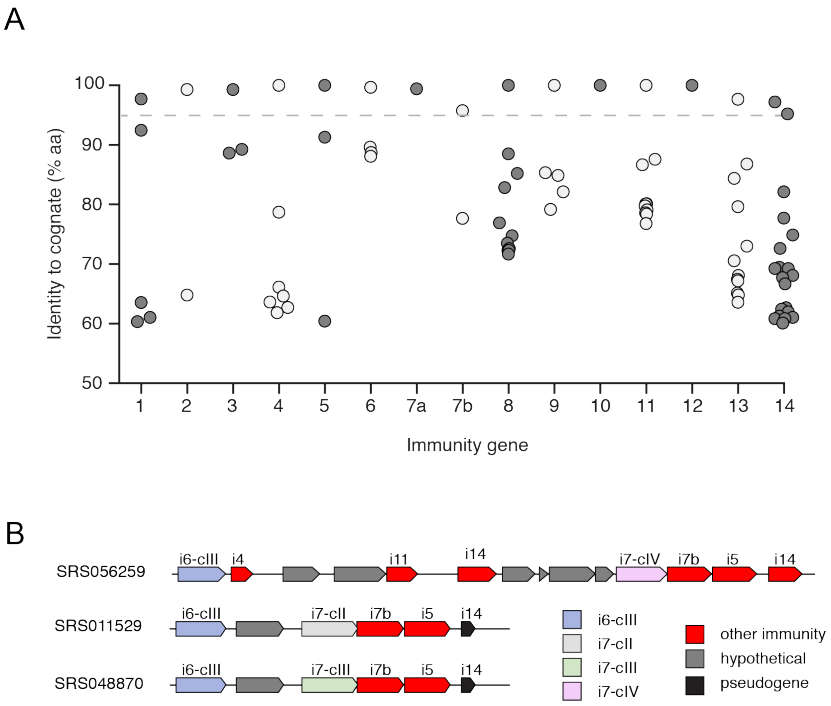
References

1. Barrangou, R., et al., *CRISPR provides acquired resistance against viruses in prokaryotes.* Science, 2007. 315(5819): p. 1709-12.
2. Cornforth, D.M. and K.R. Foster, *Competition sensing: the social side of bacterial stress responses.* Nat Rev Microbiol, 2013. 11(4): p. 285-93.
3. Hille, F., et al., *The Biology of CRISPR-Cas: Backward and Forward.* Cell, 2018. 172(6): p. 1239-1259.
4. Sender, R., S. Fuchs, and R. Milo, *Revised Estimates for the Number of Human and Bacteria Cells in the Body.* PLoS Biol, 2016. 14(8): p. e1002533.
5. Verster, A.J., et al., *The Landscape of Type VI Secretion across Human Gut Microbiomes Reveals Its Role in Community Composition.* Cell Host Microbe, 2017. 22(3): p. 411-419 e4.
6. Wexler, A.G., et al., *Human symbionts inject and neutralize antibacterial toxins to persist in the gut.* Proc Natl Acad Sci U S A, 2016. 113(13): p. 3639-44.
7. Hecht, A.L., et al., *Strain competition restricts colonization of an enteric pathogen and prevents colitis.* EMBO Rep, 2016. 17(9): p. 1281-91.
8. Chatzidaki-Livanis, M., N. Geva-Zatorsky, and L.E. Comstock, *Bacteroides fragilis type VI secretion systems use novel effector and immunity proteins to antagonize human gut Bacteroidales species.* Proc Natl Acad Sci U S A, 2016. 113(13): p. 3627-32.
9. Lloyd-Price, J., et al., *Strains, functions and dynamics in the expanded Human Microbiome Project.* Nature, 2017. 550(7674): p. 61-66.
10. Qin, J., et al., *A human gut microbial gene catalogue established by metagenomic sequencing.* Nature, 2010. 464(7285): p. 59-65.
11. Human Microbiome Project, C., *Structure, function and diversity of the healthy human microbiome.* Nature, 2012. 486(7402): p. 207-14.
12. Coyne, M.J., K.G. Roelofs, and L.E. Comstock, *Type VI secretion systems of human gut Bacteroidales segregate into three genetic architectures, two of which are contained on mobile genetic elements.* BMC Genomics, 2016. 17: p. 58.
13. Russell, A.B., et al., *A widespread bacterial type VI secretion effector superfamily identified using a heuristic approach.* Cell Host Microbe, 2012. 11(5): p. 538-49.
14. Ting, S.Y., et al., *Bifunctional Immunity Proteins Protect Bacteria against FtsZ-Targeting ADP-Ribosylating Toxins.* Cell, 2018.
15. Zhang, D., et al., *Polymorphic toxin systems: Comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics.* Biol Direct, 2012. 7: p. 18.
16. Steele, M.I., et al., *Diversification of Type VI Secretion System Toxins Reveals Ancient Antagonism among Bee Gut Microbes.* MBio, 2017. 8(6).
17. Kirchberger, P.C., et al., *Sequential displacement of Type VI Secretion System effector genes leads to evolution of diverse immunity gene arrays in Vibrio cholerae.* Sci Rep, 2017. 7: p. 45133.
18. Wozniak, R.A. and M.K. Waldor, *Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow.* Nat Rev Microbiol, 2010. 8(8): p. 552-63.
19. Li, J., et al., *An integrated catalog of reference genes in the human gut microbiome.* Nat Biotechnol, 2014. 32(8): p. 834-41.
20. Siguier, P., E. Gourbeyre, and M. Chandler, *Bacterial insertion sequences: their genomic impact and diversity.* FEMS Microbiol Rev, 2014. 38(5): p. 865-91.
21. Yassour, M., et al., *Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability.* Science Translational Medicine, 2016. 8(343): p. 343ra81.
22. Zhao, S., et al., *Adaptive evolution within the gut microbiome of individual people.* bioRxiv, 2017.
23. Castillo, F., A. Benmohamed, and G. Szatmari, *Xer Site Specific Recombination: Double and Single Recombinase Systems.* Front Microbiol, 2017. 8: p. 453.
24. Mazel, D., *Integrons: agents of bacterial evolution.* Nat Rev Microbiol, 2006. 4(8): p. 608-20.
25. Foster, K.R. and T. Bell, *Competition, not cooperation, dominates interactions among culturable microbial species.* Curr Biol, 2012. 22(19): p. 1845-50.

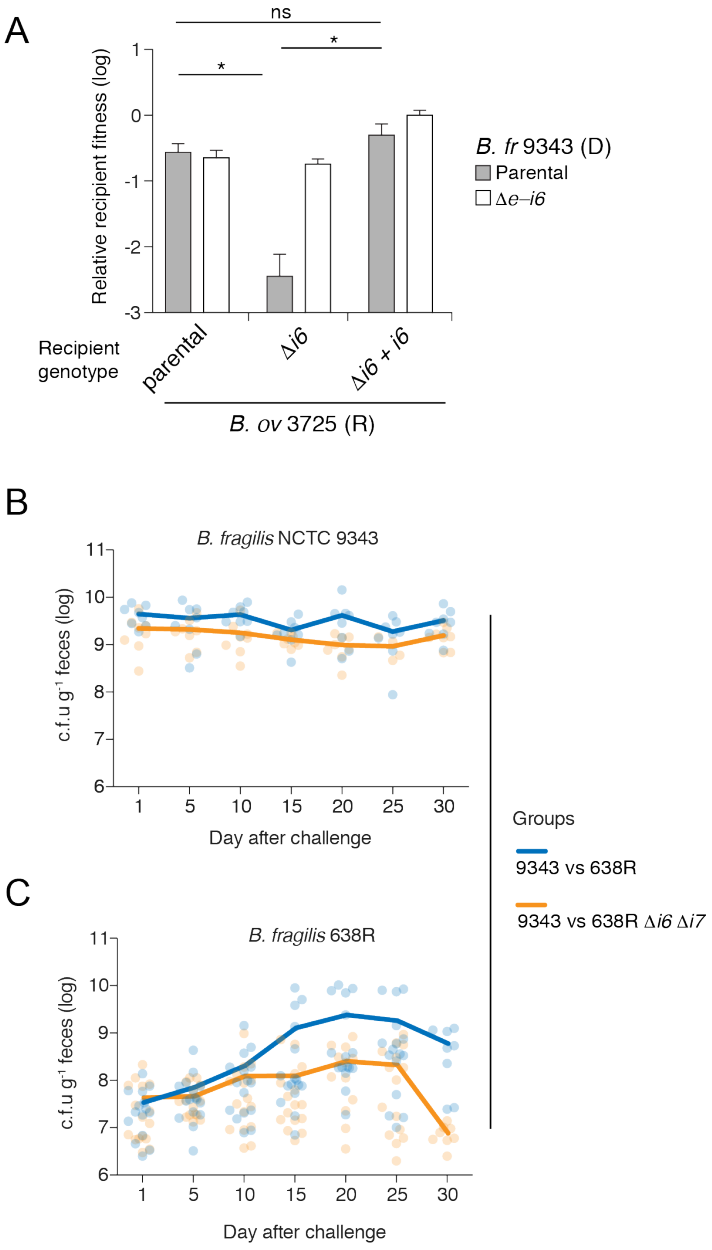
26. Poole, S.J., et al., *Identification of functional toxin/immunity genes linked to contact-dependent growth inhibition (CDI) and rearrangement hotspot (Rhs) systems.* PLoS Genet, 2011. 7(8): p. e1002217.
27. Drider, D., et al., *The continuing story of class IIa bacteriocins.* Microbiol Mol Biol Rev, 2006. 70(2): p. 564-82.
28. Coyte, K.Z., J. Schluter, and K.R. Foster, *The ecology of the microbiome: Networks, competition, and stability.* Science, 2015. 350(6261): p. 663-6.
29. Russell, A.B., et al., *A type VI secretion-related pathway in Bacteroidetes mediates interbacterial antagonism.* Cell Host Microbe, 2014. 16(2): p. 227-36.
30. Speare, L., et al., *Bacterial symbionts use a type VI secretion system to eliminate competitors in their natural host.* Proceedings of the National Academy of Sciences, 2018.
31. Bailey, T.L., et al., *MEME SUITE: tools for motif discovery and searching.* Nucleic Acids Res, 2009. 37(Web Server issue): p. W202-8.
32. Qin, J., et al., *A metagenome-wide association study of gut microbiota in type 2 diabetes.* Nature, 2012. 490(7418): p. 55-60.
33. Truong, D.T., et al., *MetaPhlAn2 for enhanced metagenomic taxonomic profiling.* Nat Methods, 2015. 12(10): p. 902-3.
34. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nature Methods, 2012. 9: p. 357.
35. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. 25(16): p. 2078-9.
36. Luo, R., et al., *SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler.* Gigascience, 2012. 1(1): p. 18.
37. Besemer, J., A. Lomsadze, and M. Borodovsky, *GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions.* Nucleic Acids Res, 2001. 29(12): p. 2607-18.
38. Bacic, M.K. and C.J. Smith, *Laboratory maintenance and cultivation of bacteroides species.* Curr Protoc Microbiol, 2008. Chapter 13: p. Unit 13C 1.
39. Koropatkin, N.M., et al., *Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices.* Structure, 2008. 16(7): p. 1105-15.
40. Degnan, P.H., et al., *Human gut microbes use multiple transporters to distinguish vitamin B(1)(2) analogs and compete in the gut.* Cell Host Microbe, 2014. 15(1): p. 47-57.
41. Hoffman, L.R., et al., *Escherichia coli dysbiosis correlates with gastrointestinal dysfunction in children with cystic fibrosis.* Clin Infect Dis, 2014. 58(3): p. 396-9.
42. Manor, O., et al., *Metagenomic evidence for taxonomic dysbiosis and functional imbalance in the gastrointestinal tracts of children with cystic fibrosis.* Sci Rep, 2016. 6: p. 22493.
43. Bankevich, A., et al., *SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing.* J Comput Biol, 2012. 19(5): p. 455-77.
44. Seemann, T., *Prokka: rapid prokaryotic genome annotation.* Bioinformatics, 2014. 30(14): p. 2068-9.
45. Koren, S., et al., *Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation.* Genome Res, 2017. 27(5): p. 722-736.
46. Martens, E.C., H.C. Chiang, and J.I. Gordon, *Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont.* Cell Host Microbe, 2008. 4(5): p. 447-57.
47. Paik, J., et al., *Potential for using a hermetically-sealed, positive-pressured isocage system for studies involving germ-free mice outside a flexible-film isolator.* Gut Microbes, 2015. 6(4): p. 255-65.
48. Silverman, J.M., et al., *Haemolysin coregulated protein is an exported receptor and chaperone of type VI secretion substrates.* Mol Cell, 2013. 51(5): p. 584-93.
49. Cardona, S.T. and M.A. Valvano, *An expression vector containing a rhamnose-inducible promoter provides tightly regulated gene expression in Burkholderia cenocepacia.* Plasmid, 2005. 54(3): p. 219-28.



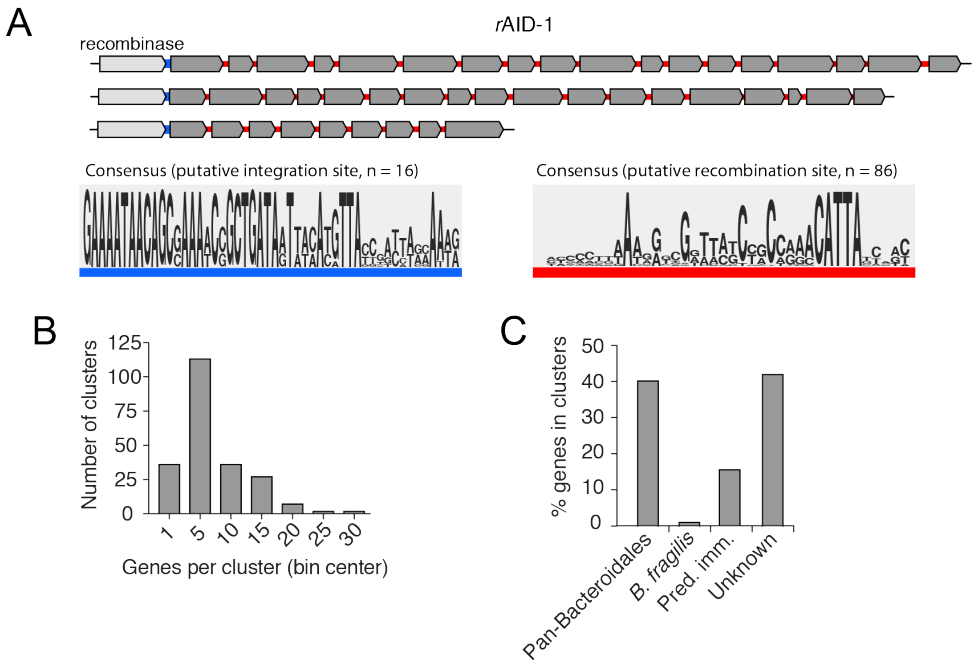
Supplemental Fig. 1: Species association and genomic context of orphan immunity genes in the genus *Bacteroides*. A-B. Mean square error (MSE) values from an analysis of the correlation between abundance of orphan immunity genes of the indicated clades (i6, A; i7, B) and the abundance of the noted species of *Bacteroides;* top six most well-correlated species are shown. The species with the lowest MSE value are plotted in Fig. 2c-g. **C**. To scale schematic depicting an integrative and conjugative element from *B. fragilis* 638R harboring the AID-1 cluster depicted in Fig. 2h.



Supplemental Fig. 2: Diversity and genomic context of orphan immunity genes in human gut microbiomes. **A.** Percent amino acid identity of unique genes homologous to *B. fragilis* T6SS cognate immunity identified through BLAST analysis of the Integrated Gene Catalog (IGC) (minimum E-value, 10^{-40}, minimum percent identity, 60%) [19]. Dashed line indicates 95% identity, denoting a conservative cutoff for cognate versus non-cognate immunity proteins. **B**. To scale depiction of metagenome-assembled scaffolds harboring orphan immunity genes from the Human Microbiome Project [9, 11]. Colors indicate predicted gene functions as in Fig. 2.



Supplemental Fig. 3: Orphan immunity genes enhance the fitness of *Bacteroides* strains *in vitro* and *in vivo*. **A.** T6SS-dependent competitiveness of parental strains of *B. ovatus* 3725 D1 iv bearing in-frame deletions of indicated orphan immunity genes or complemented with i6 constitutively expressed from a neutral chromosomal site, during *in vitro* growth competition experiments with *B. fragilis* NCTC 9343 parental strains or strains bearing in-frame deletions as indicated. Competitive index was determined by comparison of final and initial c.f.u. ratios of indicated strains (recipient to donor ratio), normalized to the corresponding competitions with *B. fragilis* NCTC 9343 lacking *tssC* (T6-inactive). Data represent mean ± s.d. for three technical replicates from each of at least three biological replicates. Asterisks indicate statistically significant differences between indicated mean values (t-test, $p < 0.01$). **B-C.** Recovery of *B. fragilis* NCTC 9343 (**B**) or *B. fragilis* 638R (**C**) colony forming units (c.f.u. per gram feces) from pairwise competitions between the indicated *B. fragilis* strains in germ-free mice (blue, parental NCTC 9343 vs. parental 638R; orange, parental NCTC 9343 vs. 638R lacking orphan immunity). Data presented is that included in Fig. 3C.



Supplemental Fig. 4: Conserved features characteristic of *r*AID-1 systems. A. Motif enrichment analysis from 16 intergenic sequences encompassing the region immediately 3' of the *r*AID-1 recombinase stop codon to the start codon of the first downstream open reading frame as identified by MEME [31] (left, blue). Motif enrichment analysis from 86 intergenic sequences (red) between the ORFs of 6 *r*AID-1 clusters (*B. fragilis* NCTC 9343, *B. cellulosilyticus* WH2, *B. ovatus* 3725 D1 iv, *Paraprevotella clara* YIT 11840, *Parabacteroides goldsteinii* dnLKV18, and *Parabacteroides gordonii* MS-1), as identified by MEME [31] (right, red). **B.** Frequency distribution of gene number in *r*AID-1 clusters (n = 1247 genes in 226 clusters). Bin width is 5 genes. **C.** Composition of genes in *r*AID-1 clusters (n = 226 clusters) as determined by profile HMM scans and BLAST analysis against a curated database of Bacteroidales T6SS immunity genes [5, 15].