Human gut bacteria contain acquired interbacterial defence systems

https://doi.org/10.1038/s41586-019-1708-z

Received: 22 August 2018

Accepted: 20 September 2019

Published online: 30 October 2019

Benjamin D. Ross[1,12], Adrian J. Verster[2,12], Matthew C. Radey[1], Danica T. Schmidtke[1], Christopher E. Pope[3], Lucas R. Hoffman[1,3,4], Adeline M. Hajjar[5], S. Brook Peterson[1], Elhanan Borenstein[2,6,7,8,9,13]* & Joseph D. Mougous[1,10,11,13]*

The human gastrointestinal tract consists of a dense and diverse microbial community, the composition of which is intimately linked to health. Extrinsic factors such as diet and host immunity are insufficient to explain the constituents of this community, and direct interactions between co-resident microorganisms have been implicated as important drivers of microbiome composition. The genomes of bacteria derived from the gut microbiome contain several pathways that mediate contact-dependent interbacterial antagonism[1–3]. Many members of the Gram-negative order Bacteroidales encode the type VI secretion system (T6SS), which facilitates the delivery of toxic effector proteins into adjacent cells[4,5]. Here we report the occurrence of acquired interbacterial defence (AID) gene clusters in Bacteroidales species that reside within the human gut microbiome. These clusters encode arrays of immunity genes that protect against T6SS-mediated intra- and inter-species bacterial antagonism. Moreover, the clusters reside on mobile elements, and we show that their transfer is sufficient to confer resistance to toxins in vitro and in gnotobiotic mice. Finally, we identify and validate the protective capability of a recombinase-associated AID subtype (rAID-1) that is present broadly in Bacteroidales genomes. These rAID-1 gene clusters have a structure suggestive of active gene acquisition and include predicted immunity factors of toxins derived from diverse organisms. Our data suggest that neutralization of contact-dependent interbacterial antagonism by AID systems helps to shape 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, such as CRISPR–Cas, represent adaptations that are unique to the specific threats encountered by a bacterial lineage[6,7]. In densely colonized habitats such as the mammalian gut, overcoming contact-dependent interbacterial antagonism is a major hurdle to survival. The T6SS is a pathway widely used by gut bacteria to mediate the delivery of toxic effector proteins into neighbouring cells[8–10]. Although kin cells are innately resistant to these effectors via cognate immunity proteins, whether non-self cells in the gut can escape intoxication is unknown. Notably, several recent studies have reported that bacteria from a range of environments can encode T6S immunity genes that lack an accompanying cognate effector gene[2,9,11–13]. It has been suggested that these genes are involved in interbacterial defence, but they have yet to be functionally investigated in a physiological context.

B. fragilis T6S immunity in the human gut microbiome

To identify potential mechanisms of defence against T6S-delivered interbacterial effectors in the human gut, we mined a large collection of shotgun metagenomic samples ($n = 553$) from several studies of the human gut microbiome for sequences homologous to known immunity genes[14,15] (Supplementary Table 1). We first focused our efforts on *Bacteroides fragilis*, which has a well-described and diverse repertoire of effector and cognate immunity genes[3,8,9] (Supplementary Table 2). As expected for genes predicted to reside within the *B. fragilis* genome, sequences mapping to these immunity loci were detected at a similar abundance to that of *B. fragilis* species-specific marker genes in many microbiome samples (Fig. 1a, grey; Supplementary Table 1). However, in a second subset of samples, immunity genes were detected at a significantly higher (more than ten times) abundance than expected given the abundance of *B. fragilis* (Fig. 1a, blue). Finally, we identified a third subset of samples in which the

[1]Department of Microbiology, University of Washington, Seattle, WA, USA. [2]Department of Genome Sciences, University of Washington, Seattle, WA, USA. [3]Department of Pediatrics, University of Washington, Seattle, WA, USA. [4]Seattle Children's Hospital, Seattle, WA, USA. [5]Department of Comparative Medicine, University of Washington, Seattle, WA, USA. [6]Department of Computer Science and Engineering, University of Washington, Seattle, WA, USA. [7]Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, Israel. [8]Department of Clinical Microbiology and Immunology, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel. [9]Santa Fe Institute, Santa Fe, NM, USA. [10]Department of Biochemistry, University of Washington, Seattle, WA, USA. [11]Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA. [12]These authors contributed equally: Benjamin D. Ross, Adrian J. Verster. [13]These authors jointly supervised this work: Elhanan Borenstein and Joseph D. Mougous *e-mail: elbo@uw.edu; mougous@uw.edu

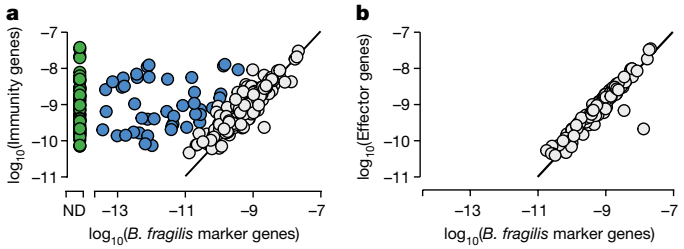


Fig. 1 | T6SS orphan immunity genes are found in human gut microbiomes. **a**, **b**, Comparison of abundance of *B. fragilis*-specific T6SS immunity genes (**a**) or effector genes (**b**) with *B. fragilis* marker genes in adult microbiome samples derived from the Human Microbiome Project (HMP) and METAgenomics of the Human Intestinal Tract (MetaHIT) studies (Supplementary Table 1). Abundance values denote the number of reads mapped to the gene, normalized by gene length and total number of reads in the sample. Abundances are calculated as the average abundance of all immunity, effector or *B. fragilis* species-specific marker genes. Samples with undetectable *B. fragilis* (green) and samples in which immunity gene abundance exceeds that of *Bacteroides* by over tenfold (blue) are highlighted. ND, not detected.

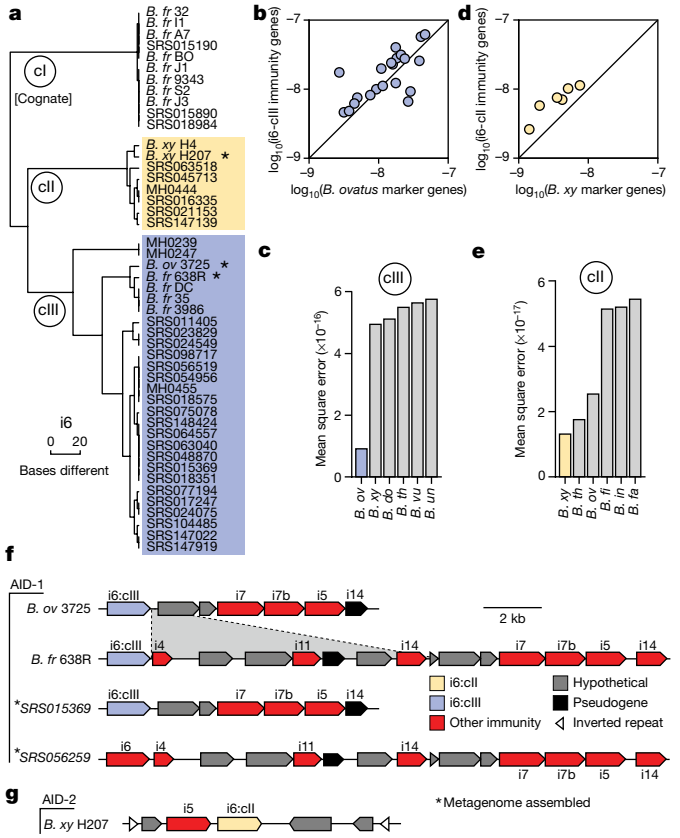


Fig. 2 | T6SS orphan immunity gene clusters are encoded by several species in the human gut microbiome. **a**, Dendrogram depicting hierarchical clustering of orphan immunity gene i6 sequences extracted from genomes ($n = 15$) and metagenomes ($n = 32$) derived from the indicated HMP (SRS) or MetaHIT (MH) samples. Sequence clades are denoted cI–III. Asterisks indicate strains shown in **f** and **g**. **b**–**d**, Comparison of abundance of genes from the indicated immunity clades (colours as in **a**) with marker genes from *B. ovatus* (**b**) or *B. xylanisolvens* (**d**) in adult microbiome samples (Supplementary Table 1). **c**, **e**, Linear model error values for the six species best fitting i6:cIII (**c**) and i6:cII (**e**) gene abundances calculated as in Fig. 1. **f**, **g**, Representative AID-1 (**f**) and AID-2 (**g**) gene clusters containing homologues of the indicated *B. fragilis* T6S immunity genes. The i6 gene of SRS056259 did not meet our sequence depth coverage requirements for inclusion in i6:cIII. The *B. xylanisolvens* strain in **g** was sequenced as a part of this study (BioProject PRJNA484981). A region of difference between *B. ovatus* 3725 and *B. fragilis* 638R clusters is highlighted. All strain abbreviations are defined in Supplementary Table 3.

sequences of immunity genes were detected in the absence of *B. fragilis* (Fig. 1a, green). These latter sequences include close homologues of 12 out of 14 unique immunity genes (i1–i14) encoded by *B. fragilis* (Extended Data Fig. 1a). In contrast to the pattern observed for immunity genes, we found no samples in which the abundance of corresponding cognate effector genes markedly exceeded that of *B. fragilis* (Fig. 1b, Supplementary Table 1).

Orphan immunity genes are encoded by many species

The detection of *B. fragilis* immunity gene homologues in samples in which we were unable to detect *B. fragilis* strongly suggests that these elements are encoded by other bacteria in the gut. Indeed, BLAST searches revealed the genomes of several *Bacteroides* spp. that contain *B. fragilis* T6S immunity gene homologues, including *B. ovatus* (i6, i7, i5 and i14), *B. vulgatus* (i8 and i13), *B. helcogenes* (i1, i9 and i10) and *B. coprocola* (i8 and i13). To determine whether these bacteria could also account for the presence of immunity genes in the human gut microbiome, we assembled full-length predicted immunity genes from the metagenomic sequencing reads of individual microbiomes. We limited this assembly to homologues of i6—the most prevalent immunity gene detected in samples lacking *B. fragilis* (Extended Data Fig. 1a). Clustering of the recovered homologues showed that most i6 sequences distribute into three discrete clades that differ by several nucleotide substitutions (i6:cI, cII and cIII) (Fig. 2a and Supplementary Table 3). A comparison of these immunity sequences to available bacterial genomes revealed a clade matching cognate immunity genes in *B. fragilis* (i6:cI). In addition, we found an i6 sequence homologous to i6:cIII in the genome of *B. ovatus*, which we previously found does not contain the cognate T6SS effector gene[8].

To identify the species that encode these sequences in human gut microbiomes, we used a simplified linear model to identify *Bacteroides* spp., the abundance of which in microbiome samples best fits that of each immunity sequence clade. We found that i6:cIII is best explained in gut metagenomes by *B. ovatus* (Fig. 2b, c), which suggests that although reference genomes of both *B. fragilis* and *B. ovatus* contain these sequences, it is most often contained by the latter in natural populations. We could not confidently define a single species containing i6:cII by this method (Fig. 2d, e); therefore, we applied the same analysis pipeline to an infant microbiome dataset for which matching stool samples were available[16] (Extended Data Fig. 1b and Supplementary Table 4). Whole-genome sequencing of isolates identified *B. xylanisolvens* as a bacterium containing i6:cII in these samples. Notably, this species best fit the abundance of i6:cII genes in the large adult metagenomic datasets, albeit by a narrow margin (Fig. 2d). On the basis of these observations, we hypothesized that orphan immunity genes—encoded by *B. fragilis* and other species of *Bacteroides*—have an adaptive role in the gut by providing defence during intra- and inter-species antagonism.

AID system immunity genes neutralize T6S toxins

To gain insight into the function of orphan immunity genes, we examined their genomic context in available reference genomes and assembled sequence scaffolds from metagenomic data. We found that homologues of *B. fragilis* T6S immunity genes i6 and i7 are located together within discrete gene clusters, which we termed AID-1 (acquired interbacterial defence 1) systems, in several *Bacteroides* strains and in microbiome samples (Fig. 2f). Within the AID-1 gene cluster, we identified distant homologues and pseudogenized remnants of additional *B. fragilis* immunity genes, including i4, i5, i11 and i14 (Fig. 2f and Supplementary Table 5). These findings prompted us to search for more distant homologues of *B. fragilis* T6S immunity genes. This revealed

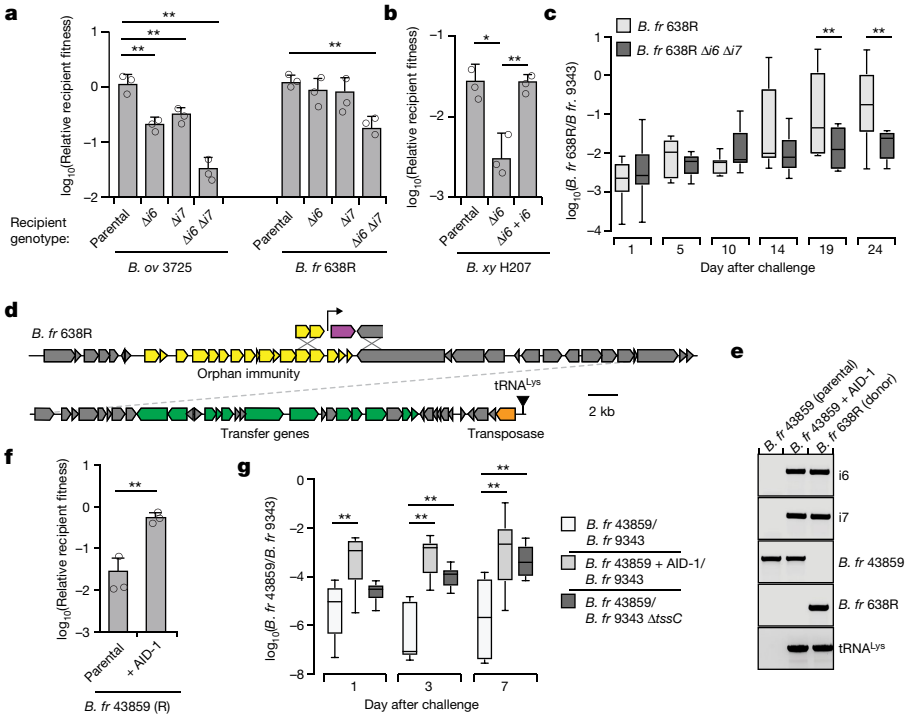


Fig. 3 | Orphan immunity genes are mobile and protect against T6S-delivered toxins. a, **b**, Outcomes of growth competition assays between the indicated strains containing AID-1 (**a**) or AID-2 (**b**) versus *B. fragilis* 9343. Relative recipient fitness was determined by calculating the ratio of final to initial colony-forming units (c.f.u.) and normalizing to corresponding experiments with T6S-inactive *B. fragilis* 9343 (Δ*tssC*). Data are mean ± s.d. of three technical replicates indicative of at least three biological replicates. $^{*}P < 0.05$, $^{**}P < 0.01$, unpaired two-tailed *t*-test. **c**, Outcome of pairwise competition between the indicated *B. fragilis* strains in germ-free mice. Mice were colonized with *B. fragilis* 9343 for one week and challenged with 638R ($n = 8$ mice per group for each of two independent experiments). For box plots, the middle line denotes the mean for each group; the box denotes the interquartile range; and the whiskers denote the minimum and maximum values. $^{**}P < 0.01$, Mann–Whitney *U*-test for each time point. **d**, Schematic depicting a *B. fragilis* 638R ICE containing the AID-1 cluster depicted in Fig. 2f. **e**, PCR analysis of an AID-1 transfer experiment. Schematic with primer locations provided in Extended Data Fig. 3f. Transfer data are representative of two independent biological replicates. **f**, Outcomes of growth competition assays between *B. fragilis* 43859 or a derivative AID-1 ICE recipient and *B. fragilis* 9343. Data presentation and statistics are as in **a** and **b**. **g**, Results of pairwise competition between the indicated *B. fragilis* strains in germ-free mice ($n = 6$ mice from two independent experiments). Statistics and data presentation are as in **c**.

gene clusters containing orphan homologues of i1, i3, i8, i9, i10 and i13 in diverse *Bacteroides* genomes and we also identified distant homologues of these genes in a metagenomic gene catalogue (Extended Data Fig. 2a, b). In *B. xylanisolvens*, we found that genes belonging to i6:cII are located in a unique, but analogous context adjacent to a homologue of i5 on an apparent transposable element[17] (Fig. 2g). We designated this sequence AID-2.

We next defined the phenotypic implications of orphan immunity genes of *Bacteroides* spp. during interbacterial competition. *B. fragilis* 9343 encodes the cognate effectors for i6 and i7, and previous data demonstrate that the corresponding toxins antagonize assorted *Bacteroides* spp. in vitro and in gnotobiotic mice[9]. We thus used this strain in growth competition assays against *Bacteroides* spp. bearing orphan immunity genes, derivative strains containing deletions of these genes, or genetically complemented strains. These experiments showed that in both *B. ovatus* and *B. fragilis*, AID-1 system genes grant immunity against corresponding T6S effectors (Fig. 3a, Extended Data Fig. 3a, b). The i6 and i7 genes of *B. ovatus* did not influence the outcome of its competition with *B. fragilis* 638R, which possesses an orthogonal effector repertoire (Extended Data Fig. 3c). Finally, we also found that an i6:cII gene from a *B. xylanisolvens* AID-2 system gives this bacterium protection against e6 of *B. fragilis* 9343 (Fig. 3b). Together, these data show that the orphan immunity genes of several *Bacteroides* species—localized to AID systems—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 owing to active strain exclusion via the T6SS[8,18]. However, in colonization experiments in gnotobiotic mice, certain *B. fragilis* strain pairs inexplicably coexist[10]. We noted that one such pair corresponds to *B. fragilis* 9343 and *B. fragilis* 638R, the latter of which contains an AID-1 system containing homologues of i6 and i7. 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 after pre-colonization of germ-free mice with *B. fragilis* 9343 (Fig. 3c, Extended Data Fig. 3d, e). Our results indicated that the cumulative protection afforded by orphan i6 and i7 genes underlies the ability of *B. fragilis* 638R to persist during T6S-mediated antagonism in vivo.

AID-1 transfer confers protection against T6S

Notably, we found that AID-1 resides on a predicted mobile integrative and conjugative element (ICE), which provides a possible explanation for its distribution[19] (Fig. 3d). To test whether this element can be transferred between strains, we performed mobilization studies using *B. fragilis* 638R as a donor and *B. fragilis* 43859 as a recipient. An antibiotic-resistance marker was inserted within AID-1 to facilitate the detection of its transfer. With this tool, we readily detected AID-1 transfer (Fig. 3e). This occurred at a frequency of approximately 5×10^{-6}, in line with previous quantification of ICE mobility in *Bacteroides* spp.[20]. Next, we asked whether the transfer of AID-1 to *B. fragilis* 43859 is sufficient to confer resistance to T6S-mediated antagonism. In vitro growth competition assays against *B. fragilis* 9343 showed that AID-1 effectively neutralizes intoxication by e6 and e7 (Fig. 3f). The receipt of AID-1 also

granted notable protection to *B. fragilis* 43859 against T6S-mediated killing in germ-free mice pre-colonized with *B. fragilis* 9343 (Fig. 3g). Together, these findings indicate that the transfer of a mobile orphan immunity island to a naive *Bacteroides* strain is sufficient to provide defence against T6S effectors.

Deciphering the contribution of individual gene products, or even whole pathways, to bacterial fitness in complex microbial communities is challenging. We reasoned that the identification of orphan immunity genes in human gut metagenomes, coupled with our ability to infer their organismal source, provided an opportunity to measure the effect 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. We found that the average abundance of *B. ovatus* strains with orphan immunity genes significantly exceeds that of those without orphan genes (Extended Data Fig. 3g). One interpretation of this finding is that the acquisition of i6 and i7 allows *B. ovatus* to increase its niche; however, there are several potential caveats inherent to these correlative data that cannot be ruled out. For example, *B. ovatus* strains that contain i6 and i7 might be related and enriched for other fitness determinants that account for their abundance.

Recombinase-associated AID systems are prevalent

Given the benefit of orphan immunity genes against *B. fragilis* effectors, we hypothesized that this mechanism of inhibiting interbacterial antagonism should extend to effectors produced by other species. We previously found that *B. fragilis* is antagonized by other *Bacteroides* spp. in the human gut microbiome[8]. In addition to the T6SS present exclusively in *B. fragilis*, this species and other *Bacteroides* species can possess other T6SSs, referred to as GA1 (genetic architecture 1) and GA2, with a distinct and non-overlapping repertoire of effector and immunity genes[3]. The effectors of these systems exhibit hallmarks of antibacterial toxins and we demonstrated their ability to mediate interbacterial antagonism[2,8] (Fig. 4, Extended Data Fig. 4). Therefore, we searched *B. fragilis* genomes for sequences that are homologous to the immunity genes corresponding to these systems. In 29 out of the 122 available *B. fragilis* genomes, we identified apparent orphan homologues of these immunity genes grouped within gene clusters (Fig. 4a). Although analogous to the AID-1 and AID-2 systems, these clusters have several unique characteristics including skewed GC content, conservation of a gene encoding a predicted XerD-family tyrosine recombinase, and repetitive intergenic sequences[21] (Extended Data Fig. 5a–c). These often-large gene clusters, hereafter referred to as recombinase-associated AID (rAID-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[2] (Fig. 4a, Extended Data Fig. 5d, e, Supplementary Table 6). Using the shared characteristics of *B. fragilis* rAID-1 systems, we searched for related gene clusters across sequenced Bacteroidales genomes. We found that more than half of sequenced bacteria belonging to this order possess a rAID-1 system (226 out of 423) (Fig. 4b, Supplementary Tables 6, 7). In summary, these gene clusters contain 579 unique genes, which encompass homologues of 25 Bacteroidales T6S immunity genes.

The prevalence of rAID-1 genes in Bacteroidales genomes suggested that these elements may be common in the human gut microbiome. To investigate this, we searched metagenomic data for sequences that map to Bacteroidales T6S orphan immunity genes found within rAID-1 systems. Notably, we found one or more rAID-1 immunity genes in 551 out of 553 samples using a 97% sequence identity threshold to map reads (Supplementary Table 8). These rAID-1 immunity genes diverge considerably from corresponding cognate immunity genes (corresponding to 32–91% amino acid identity), which suggests that the latter are an unlikely source of significant false positives in this analysis. We also searched the same samples for rAID-1-associated recombinase

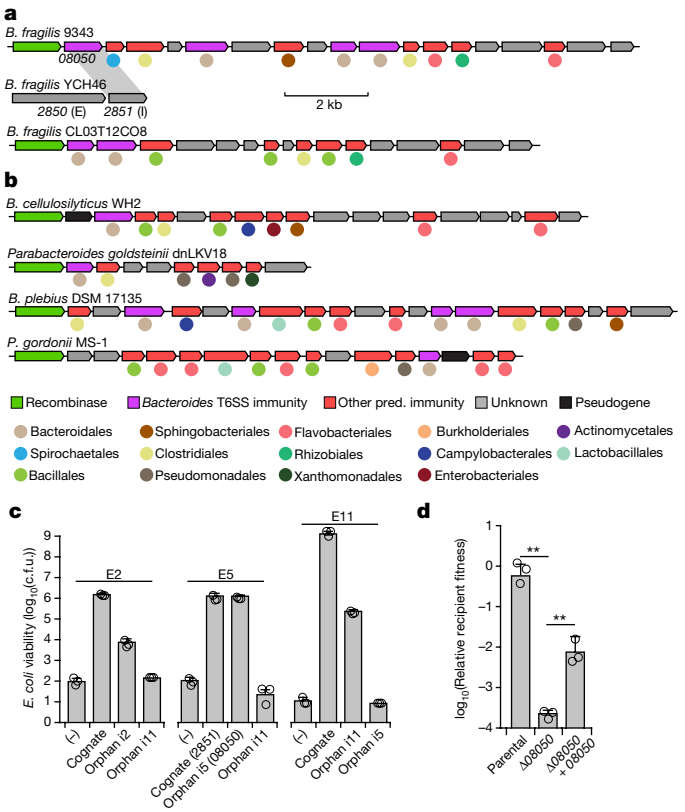


Fig. 4 | rAID systems encode toxin-neutralizing immunity genes and are prevalent in human gut microbiomes. a, **b**, rAID-1 clusters from the indicated *B. fragilis* (**a**) or Bacteroidales (**b**) species. rAID-1 genes were assigned to functional immunity classes (indicated by gene colouring) via profile HMM scans and BLAST against a curated database of Bacteroidales T6SS immunity genes[2,8]. Coloured circles indicate taxonomic association of the top non-rAID-1 homologue. Homology (70% amino acid identity) between gene 08050 of the *B. fragilis* 9343 rAID-1 cluster and a T6S cognate immunity gene from *B. fragilis* YCH46 (2851) is indicated. **c**, Viable *E. coli* cells recovered from cultures expressing the indicated proteins (see Supplementary Table 10 for locus tags). Data are mean ± s.d. of three technical replicates representative of three independent biological replicates. **d**, Outcomes of growth competition assays between the indicated *Bacteroides* strains ($n = 3$ biologically independent samples). The relevant rAID-1 gene of *B. fragilis* 9343 and its corresponding effector within *B. fragilis* YCH46 are depicted in **a**. $^{**}P < 0.01$, unpaired two-tailed *t*-test. Data are mean ± s.d.

sequences. Although recombinase genes are widely distributed across bacteria, close homologues (more than 50% amino acid identity) of those found associated with rAID-1 systems are restricted to this context and only found in Bacteroidales genomes. Consistent with this, we found that the abundance of rAID-1 recombinase genes correlates strongly with the genus *Bacteroides* (Extended Data Fig. 5f).

Divergent rAID-1 immunity genes protect against T6S

Orphan immunity genes encoded within rAID-1 clusters diverge more from cognate immunity than do those within AID-systems. Thus, we sought to experimentally validate the ability of rAID-1 immunity genes to protect bacteria from intoxication. Because most bacteria containing rAID-1 systems have limited genetic tools, we used *Escherichia coli* to identify three Bacteroidales T6S effector genes that intoxicate cells in a manner that is neutralized by cognate immunity. In each case, we found that co-expression of these effector genes with corresponding rAID-1-associated orphan immunity genes, but not mismatched orphan immunity genes, restored *E. coli* growth (Fig. 4c). Both of the genes from

one effector–orphan immunity pair that we validated derive from genetically tractable strains: *B. fragilis* YCH46 (effector, 2850) and *B. fragilis* 9343 (orphan immunity, 08050). In vitro growth competition experiments with these strains, and mutant and genetically complemented derivatives, showed that an endogenous rAID-1 orphan immunity gene of *B. fragilis* 9343 can neutralize a T6S-delivered toxin (Fig. 4d).

The orphan immunity systems that we defined consist of many genes and their expression could incur a substantial metabolic burden. As a first step towards understanding the regulation of AID systems, we performed quantitative PCR with reverse transcription (qRT–PCR) analysis to compare the expression of the systems in the presence and absence of a competitor strain. These studies provided evidence that transcription of both systems is induced by co-cultivation with a competitor strain (Extended Data Fig. 5g). We also examined meta-transcriptomic data for evidence of AID expression. Owing to a paucity of such data available for samples definitively containing AID-1 and AID-2, we could not systematically quantify the expression of these systems. However, using conservative criteria for defining rAID-1-associated genes in meta-transcriptomic data, we found evidence for the expression of this system in every sample derived from a large study ($n = 156$)[22] (Supplementary Table 9). In some samples, such as those with high levels of *Bacteroides*, rAID-1 genes accounted for nearly 1 in 10,000 of all meta-transcriptomic reads. Together with our functional characterization of AID systems, these findings suggest that acquisition and maintenance of consolidated orphan immunity determinants is a common mechanism by which Bacteroidales defend against interbacterial antagonism in the human gut microbiome.

Discussion

Mounting evidence suggests that competitive interactions between bacteria predominate in many environments[23]. This evolutionary pressure has undoubtedly led to the wide dissemination of idiosyncratically orphaned immunity genes that are predicted to provide resistance to diverse antagonistic pathways[2,11,24,25]. Modelling studies predict that interbacterial antagonism is a crucial contributor to the maintenance of a stable gut community[26]. 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 the diversification and expansion of T6S effectors. Deciphering the link between orphan immunity genes and the bacteria harbouring the cognate effectors may help to shed light on the physical connectivity of bacteria in the gut microbiome.

It is now appreciated that phage defence mechanisms, including the adaptive system CRISPR, are crucial for bacteria to cope with the omnipresent threat and deleterious outcome of phage infection[7]. 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[2]. Our characterization of AID systems encoded by prevalent members of the human gut microbiota seems to reconcile these observations and demonstrate that the neutralization of contact-dependent interbacterial antagonism can be a critical mechanism for survival in polymicrobial environments. In addition, it suggests that analogous to the immune system of vertebrates, that of bacteria includes arms specialized in viral or bacterial defence.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1708-z.

1. Whitney, J. C. et al. A broadly distributed toxin family mediates contact-dependent antagonism between Gram-positive bacteria. *eLife* **6**, e26938 (2017).
2. Zhang, D., de Souza, R. F., Anantharaman, V., Iyer, L. M. & Aravind, L. Polymorphic toxin systems: comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. *Biol. Direct* **7**, 18 (2012).
3. Coyne, M. J., Roelofs, K. G. & Comstock, L. E. 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 (2016).
4. Russell, A. B. et al. A type VI secretion-related pathway in Bacteroidetes mediates interbacterial antagonism. *Cell Host Microbe* **16**, 227–236 (2014).
5. Hood, R. D. et al. A type VI secretion system of *Pseudomonas aeruginosa* targets a toxin to bacteria. *Cell Host Microbe* **7**, 25–37 (2010).
6. Cornforth, D. M. & Foster, K. R. Competition sensing: the social side of bacterial stress responses. *Nat. Rev. Microbiol.* **11**, 285–293 (2013).
7. Hille, F. et al. The biology of CRISPR–Cas: backward and forward. *Cell* **172**, 1239–1259 (2018).
8. Verster, A.J. et al. The landscape of type VI secretion across human gut microbiomes reveals its role in community composition. *Cell Host Microbe* **22**, 411–419 (2017).
9. Wexler, A. G. et al. Human symbionts inject and neutralize antibacterial toxins to persist in the gut. *Proc. Natl Acad. Sci. USA* **113**, 3639–3644 (2016).
10. Hecht, A. L. et al. Strain competition restricts colonization of an enteric pathogen and prevents colitis. *EMBO Rep.* **17**, 1281–1291 (2016).
11. Kirchberger, P. C., Unterweger, D., Provenzano, D., Pukatzki, S. & Boucher, Y. Sequential displacement of type VI secretion system effector genes leads to evolution of diverse immunity gene arrays in *Vibrio cholerae*. *Sci. Rep.* **7**, 45133 (2017).
12. Steele, M. I., Kwong, W. K., Whiteley, M. & Moran, N. A. Diversification of type VI secretion system toxins reveals ancient antagonism among bee gut microbes. *mBio* **8**, e26938 (2017).
13. Ting, S. Y. et al. Bifunctional immunity proteins protect bacteria against Ftsz-targeting ADP-ribosylating toxins. *Cell* **175**, 1380–1392 (2018).
14. Lloyd-Price, J. et al. Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature* **550**, 61–66 (2017).
15. Qin, J. et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
16. Manor, O. et al. Metagenomic evidence for taxonomic dysbiosis and functional imbalance in the gastrointestinal tracts of children with cystic fibrosis. *Sci. Rep.* **6**, 22493 (2016).
17. Siguier, P., Gourbeyre, E. & Chandler, M. Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiol. Rev.* **38**, 865–891 (2014).
18. Zhao, S. et al. Adaptive evolution within gut microbiomes of healthy people. *Cell Host Microbe* **25**, 656–667.e8 (2019).
19. Wozniak, R. A. & Waldor, M. K. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat. Rev. Microbiol.* **8**, 552–563 (2010).
20. Stevens, A. M., Shoemaker, N. B. & Salyers, A. A. The region of a *Bacteroides* conjugal chromosomal tetracycline resistance element which is responsible for production of plasmidlike forms from unlinked chromosomal DNA might also be involved in transfer of the element. *J. Bacteriol.* **172**, 4271–4279 (1990).
21. Castillo, F., Benmohamed, A. & Szatmari, G. Xer site specific recombination: double and single recombinase systems. *Front. Microbiol.* **8**, 453 (2017).
22. Abu-Ali, G. S. et al. Metatranscriptome of human faecal microbial communities in a cohort of adult men. *Nat. Microbiol.* **3**, 356–366 (2018).
23. Foster, K. R. & Bell, T. Competition, not cooperation, dominates interactions among culturable microbial species. *Curr. Biol.* **22**, 1845–1850 (2012).
24. 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.* **7**, e1002217 (2011).
25. Drider, D., Fimland, G., Héchard, Y., McMullen, L. M. & Prévost, H. The continuing story of class IIa bacteriocins. *Microbiol. Mol. Biol. Rev.* **70**, 564–582 (2006).
26. Coyte, K. Z., Schluter, J. & Foster, K. R. The ecology of the microbiome: networks, competition, and stability. *Science* **350**, 663–666 (2015).

Methods

Microbiome data

Metagenomic data from healthy adults were obtained from several large-scale sequencing projects. We specifically obtained 147 samples from the Human Microbiome Project (HMP) 1.0, 100 samples from HMP 1.2, and 99 and 207 samples from two different MetaHIT datasets[14,15,27,28]. We further obtained paired metagenomic–metatranscriptomic data from a study of 156 individuals[22]. Finally, we obtained a database of genes identified from 1,267 assembled metagenomes as part of the integrated gene catalogue (IGC)[29].

Analysis of gene and species abundances in microbiome samples

We previously compiled a list of T6SS immunity and effector genes[8]. We also compiled a list of species-specific marker genes for all *Bacteroides* species obtained from MetaPhlAn 2.0[30]. 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[31]. 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. For the total *Bacteroides* abundance, we used the sum of all species-specific marker genes in the genus. Samples were only included in an analysis if they had at least 10 reads mapping to the T6SS genes in question (effectors, immunity or recombinases). On the basis of the abundance of GA3 immunity genes and *B. fragilis*, we split samples into those in which *B. fragilis* was not detectable, those in which the immunity gene had more than ten times the abundance of the *B. fragilis* marker gene, and those in which such discrepancy between the abundance of immunity genes and that of *B. fragilis* was not observed. Meta-transcriptomics data were processed similarly to metagenomics data, except that abundance values were converted to a reads per kilobase of transcript per million mapped reads (RPKM) value for familiarity with canonical RNA sequencing analysis.

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[31,32]. 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 10 times coverage the base was set to the major allele. Sites with less than ten times coverage were assigned an ambiguous base. Finally, we only kept the reconstructed sequence in metagenomic samples where at least 90% of the sequence had more than ten times coverage. The number of single nucleotide polymorphisms between all immunity sequences, both from metagenomic samples and from *Bacteroides* genomes, was calculated and used to populate a distance matrix. Because obtained distances were small (for example, a single base difference), we used hierarchical clustering (with complete linkage), rather than standard phylogenetic reconstruction methods, to visualize the relatedness between different sequences. Sequence clades defined by hierarchical clustering are denoted (cI–III), as discussed in the text.

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 three 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 used 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 were assembled using SoapDeNovo2 with a kmer length of 63 and an average insert size of 200[33]. BLAST was used to identify the contig that contained the orphan immunity gene, and GeneMarkS was used to predict protein-coding genes[34].

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 (LB) agar. Unless otherwise noted, *Bacteroides* strains were cultured under anaerobic conditions on brain heart infusion (BHI) agar (Becton Dickinson) supplemented with 50 μg ml^{-1} haemin and 10% sodium bicarbonate (BHIS)[35]. Antibiotics and chemicals were added to media as needed at the following concentrations: trimethoprim 50 μg ml^{-1}, carbenicillin 150 μg ml^{-1}, gentamicin 15 μg ml^{-1} (*E. coli*), gentamicin 60 μg ml^{-1} (*Bacteroides*), erythromycin 12.5 μg ml^{-1}, tetracycline 6 μg ml^{-1}, chloramphenicol 12 μg ml^{-1}, floxuridine (FUdR) 200 μg ml^{-1}.

Genetic techniques

Standard molecular procedures were used for the 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 (Supplementary Table 10). 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[36]. In brief, 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* was confirmed by PCR. Subsequent deletion of orphan immunity genes was performed in Δ*tdk* strains via a similar counter selection strategy, except employing the suicide plasmid pExchange in place of pKNOCK[9]. Genomic deletions were confirmed by PCR. Gene complementation was performed by cloning genes into pNBU2-erm_us1311 for constitutive expression[37].

Isolation of *Bacteroides* strains from faecal samples

Faecal samples from healthy infants used for strain isolation were collected as part of a previous study approved by the Seattle Children's Hospital Institutional Review Board[16,38]. 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). Single colonies that exhibited esculin hydrolysis as indicated by the production of black pigment on BBE agar were sub-cultured in TYG broth with the addition of 60 μg ml^{-1} gentamicin until stationary phase and then were frozen at −80 °C after the addition of sterile glycerol to 20% final concentration. Single colonies isolated from these stocks were subsequently screened by PCR with primers targeting the orphan i6 gene as assembled from metagenomic short read sequence data[16].

Genome sequencing

Genomic DNA used for Illumina sequencing was prepared by collecting *Bacteroides* strains grown overnight on BHIS blood agar plates. Cells resuspended from plates were washed in PBS 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. AID clusters often appear in highly repetitive genomic contexts (for example, mobile elements) and are often split into multiple scaffolds in reference genomes. To compensate for this, we also performed long-read sequencing via PacBio on a subset of genomes. To this end, high molecular mass DNA was extracted using the Qiagen Genomic-tip Kit and sequenced by SNPsaurus using a PacBio Sequel. Hybrid long read and short read assemblies were conducted using Unicycler[39]. Species identification was performed by blast searches with species-specific marker genes[30].

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 (OD) 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 5 µl of bacterial mixtures were spotted onto pre-dried BHIS blood agar plates, in triplicate spots. Competitions were allowed to proceed for 20–24 h at 37 °C under anaerobic conditions before spots were collected into BHIS broth. Competition outcomes were quantified in one of two ways: (1) by serial dilution for enumeration of c.f.u. 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 Supplementary Table 10). For antibiotic selection, *B. fragilis* 9343 was marked with erythromycin resistance by integration of pNBU2-erm at the att1 site[37]. Other strains were either naturally tetracycline resistant, or marked by integration of pNBU2-tet-BCO1. Strains with insertions of pNBU2 were selected for matching integration sites by PCR with primers flanking att loci[40]. Interbacterial competitions between strains of *B. fragilis* occasionally exhibited T6SS-independent phenotypes that were dependent on the initial starting ratio of the strains used[41].

Interbacterial mobile element transfer assays

Allelic exchange was used to engineer a high-expression chloramphenicol resistance cassette onto the AID-1 system of *B. fragilis* 638R, replacing BF638R_2056-2058[42]. Chloramphenicol-resistant *B. fragilis* 638R cells were mixed on BHIS blood agar plates with erythromycin-resistant *B. fragilis* ATCC 43859 cells at a 1:1 ratio (OD_{600} 6.0). After overnight co-culture, bacterial mixtures were collected and plated on BHIS plates containing either erythromycin alone (to quantify c.f.u. of total ATCC cells), or erythromycin and chloramphenicol (to quantify c.f.u. of AID-1 recipient ATCC cells). Double-resistant colonies were screened individually by PCR to confirm strain identity, the presence of the AID-1 system, and the genomic integration site at a $tRNA^{Lys}$ gene (see Supplementary Table 10 for primers used).

Gnotobiotic mice studies

Germ-free 6–12-week-old female Swiss Webster mice from several litters were randomized, housed simultaneously in pairs in single Techniplast cages with a 12-h light/dark cycle, and fed a standard laboratory 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, and no statistical methods were used to determine sample size. Reasonable numbers of animals were used considering limitations of housing and maintenance under gnotobiotic conditions. *Bacteroides fragilis* strains were introduced into mice via oral gavage of 10^8 c.f.u. suspended in 0.2 ml of sterile PBS with 20% glycerol. Challenge with *B. fragilis* 638R or *B. fragilis* ATCC strains occurred 7 days after pre-colonization with *B. fragilis* 9343 strains. Colonization levels by each strain in each mouse were tracked by collection of faecal pellets over a period of 4 weeks, plating on selective BHIS agar plates (*B. fragilis* 9343 on BHIS plus erythromycin; *B. fragilis* 638R and ATCC on BHIS plus tetracycline), and subsequent absolute quantification of c.f.u. by normalization of each sample to the initial pellet weight. Differences in the strain ratio of c.f.u between groups at each time point was assessed using Mann–Whitney *U*-tests. Non-parametric tests were used following Shapiro–Wilk analysis for normality of data at each time point. Mice were confirmed to be sterile before colonization by qPCR with primers targeting the *16S* rRNA gene and free of non-*Bacteroides* contamination by plating faecal pellets on non-selective LB and BHIS plates incubated under either anaerobic or aerobic conditions[43].

Bioinformatic analysis of rAID-1 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. rAID clusters in *Bacteroidales* genomes were identified based on 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 1×10^{-100}); (ii) two or more co-directionally oriented downstream genes that 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 homologues of genes within rAID- clusters, open-reading frames (ORFs) 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 rAID clusters; therefore, these hits were discarded. The top non-rAID hit from tblastn searches with an *E*-value threshold of 1×10^{-30} was selected as the closest homologue. rAID cluster genes were assigned to interbacterial immunity gene families via hmm scans with profiles previously described[2] with an *E*-value cut-off of 1×10^{-3}. rAID cluster genes were also compared via tblastn with 46 Bacteroidales T6SS immunity genes from subtypes 1–3[3,8] with an *E*-value cut-off of 1×10^{-10}. The percentage amino acid identity with homologues was assessed if sequences could be aligned across more than 80% of their length. Motif enrichment analysis was performed on non-coding sequences within a subset of rAID-1 clusters (14 sequences immediately 3′ of the recombinase stop codon, and 86 intergenic sequences between rAID-1 ORFs), using MEME Suite 5.0.2 and default settings[44].

Heterologous expression of *Bacteroides* toxin and immunity genes

To assess the ability of cognate immunity or orphan immunity to neutralize the toxicity of a *Bacteroidales* T6SS effector, genes were cloned into *E. coli* expression vectors pScrhab2-V (effectors) and pPSV39-CV (immunity). Immunity genes were fused with the *P. aeruginosa* ribosome binding site from *hcp1* during the cloning process[45]. All cloning steps for effector genes involved growth of *E. coli* on media containing 0.1% glucose to ensure repression of expression. *E. coli* DH5a cells were co-transformed with pSchraB2 and pPSV39 plasmids bearing genes of interest. Overnight cultures were then grown from single co-transformed colonies to stationary phase in LB broth containing 50 µg ml^{-1} trimethoprim, 15 µg ml^{-1} gentamycin, and glucose. Cells were collected from these cultures and washed to remove glucose before back-dilution to an OD_{600} of 0.05 into LB broth containing 50 µg ml^{-1} trimethoprim, 15 µg ml^{-1} gentamycin, 0.05% rhamnose, and 1 mM isopropyl β-D-1-thiogalactopyranoside[45,46]. Cultures were then grown for 8 h shaking at 37 °C before

plating to allow quantification of c.f.u. Experiments were performed with technical triplicates for each of at least two biological replicates.

Gene expression analysis of AID-1 and rAID-1 systems of *B. ovatus* 3725

To assess the level of expression of genes in the AID-1 and rAID-1 systems of *B. ovatus* 3725, bacterial cells were first grown overnight on BHIS blood agar plates containing gentamycin. Cells were then resuspended in BHIS to an OD_{600} of 3.0 for *B. ovatus* monocultures, or to an OD_{600} of 0.3 for mixed co-culture experiments with *B. fragilis* 9343 at tenfold excess (OD_{600} of 3.0). Then, 5 µl volumes of bacterial mixtures were then spotted on BHIS blood agar plates. Plates were incubated at 37 °C for 2 h under anaerobic culture conditions before cells were collected directly in Buffer RLT plus β-mercaptoethanol (20 5-µl spots per condition per replicate, Qiagen RNeasy Micro Kit). Two separate rounds of DNase treatment were performed (Qiagen RNase-free DNase, Thermo Fisher Scientific Turbo DNase-free kit). RNA samples were confirmed to be free of genomic DNA by PCR with primers targeting the Bacteroides 16S rRNA gene. cDNA was generated using the High Capacity cDNA Reverse Transcription Kit (Applied Biosciences). Following synthesis, cDNA was diluted 1:10. qPCR (primers listed in Supplementary Table 10) was performed using SSO Universal SYBR Green Supermix (Bio-Rad) on a CFX96 machine (Bio-Rad). Genomic DNA was used to generated standard curves[47]. Differences in gene expression between samples were performed by normalization to the expression level of *B. ovatus* 3725 *gyrB*. Primers targeting *gyrB* were designed to target regions of the genes that are highly polymorphic between *B. fragilis* and *B. ovatus*, and species-specificity for *B. ovatus* was confirmed by PCR using *B. fragilis* genomic DNA[48].

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All data required to assess the conclusion of this research are available in the main text and Supplementary Information, have been deposited at the Sequence Read Archive (SRA) under BioProject accession PRJNA484981 or are available from https://github.com/borenstein-lab/T6SS).

Code availability

Python and R scripts used in this work are available for download (https://github.com/borenstein-lab/T6SS).

27. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
28. Qin, J. et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**, 55–60 (2012).
29. Li, J. et al. An integrated catalog of reference genes in the human gut microbiome. *Nat. Biotechnol*. **32**, 834–841 (2014).
30. Truong, D. T. et al. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat. Methods* **12**, 902–903 (2015).
31. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
32. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
33. Luo, R. et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**, 18 (2012).
34. Besemer, J., Lomsadze, A. & Borodovsky, M. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* **29**, 2607–2618 (2001).
35. Bacic, M. K. & Smith, C. J. Laboratory maintenance and cultivation of bacteroides species. *Curr. Protoc. Microbiol*. **9**, 13C.1.1–13C.1.21 (2008).
36. Koropatkin, N. M., Martens, E. C., Gordon, J. I. & Smith, T. J. Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. *Structure* **16**, 1105–1115 (2008).
37. Degnan, P. H., Barry, N. A., Mok, K. C., Taga, M. E. & Goodman, A. L. Human gut microbes use multiple transporters to distinguish vitamin B_{12} analogs and compete in the gut. *Cell Host Microbe* **15**, 47–57 (2014).
38. Hoffman, L. R. et al. *Escherichia coli* dysbiosis correlates with gastrointestinal dysfunction in children with cystic fibrosis. *Clin. Infect. Dis*. **58**, 396–399 (2014).
39. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Comput. Biol*. **13**, e1005595 (2017).
40. Martens, E. C., Chiang, H. C. & Gordon, J. I. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* **4**, 447–457 (2008).
41. García-Bayona, L. & Comstock, L. E. Bacterial antagonism in host-associated microbial communities. *Science* **361**, eaat2456 (2018).
42. Lim, B., Zimmermann, M., Barry, N. A. & Goodman, A. L. Engineered regulatory systems modulate gene expression of human commensals in the gut. *Cell* **169**, 547–558 (2017).
43. 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* **6**, 255–265 (2015).
44. Bailey, T. L. et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
45. Silverman, J. M. et al. Haemolysin coregulated protein is an exported receptor and chaperone of type VI secretion substrates. *Mol. Cell* **51**, 584–593 (2013).
46. Cardona, S. T. & Valvano, M. A. An expression vector containing a rhamnose-inducible promoter provides tightly regulated gene expression in *Burkholderia cenocepacia*. *Plasmid* **54**, 219–228 (2005).
47. Bookout, A. L., Cummins, C. L., Mangelsdorf, D. J., Pesola, J. M. & Kramer, M. F. High-throughput real-time quantitative reverse transcription PCR. *Curr. Protoc. Mol. Biol*. **73**, 15.8.1–15.8.28 (2006).
48. Caro-Quintero, A. & Ochman, H. Assessing the unseen bacterial diversity in microbial communities. *Genome Biol. Evol*. **7**, 3416–3425 (2015).

Acknowledgements We thank the UW GNAC for assistance with gnotobiotic experiments. We thank C. Sears, A. Goodman, T. Kuwahara and E. Martens for providing *Bacteroides* strains. This work was supported by National Institutes of Health (NIH) grants AI080609 (to J.D.M.), P30DK089507 (to L.R.H. as pilot study PI), R01DK095869 (to L.R.H.), K99GM129874 (to B.D.R.), R01GM124312 (to E.B.), and New Innovator Award DP2AT00780201 (to E.B.), and the Burroughs Wellcome Fund (to J.D.M.). A.J.V. was supported by a postdoctoral fellowship from the Natural Sciences and Engineering Research Council of Canada. B.D.R. was supported by a Simons Foundation-sponsored Life Sciences Research Foundation postdoctoral fellowship. E.B. is a Faculty Fellow of the Edmond J. Safra Center for Bioinformatics at Tel Aviv University. J.D.M. is an HHMI Investigator.

Author contributions B.D.R., A.J.V., A.M.H., S.B.P., E.B. and J.D.M. designed the study. B.D.R. and D.T.S. performed in vitro growth experiments; A.J.V., B.D.R. and M.C.R. performed bioinformatic analyses; B.D.R., C.E.P. and L.R.H. isolated and sequenced genomes of gut bacteria; B.D.R., D.T.S., A.M.H. and S.B.P. performed gnotobiotic mouse experiments; B.D.R., A.J.V., M.C.R., D.T.S., A.M.H., S.B.P., E.B. and J.D.M. analysed data; and B.D.R., A.J.V., S.B.P., E.B. and J.D.M. wrote the manuscript.

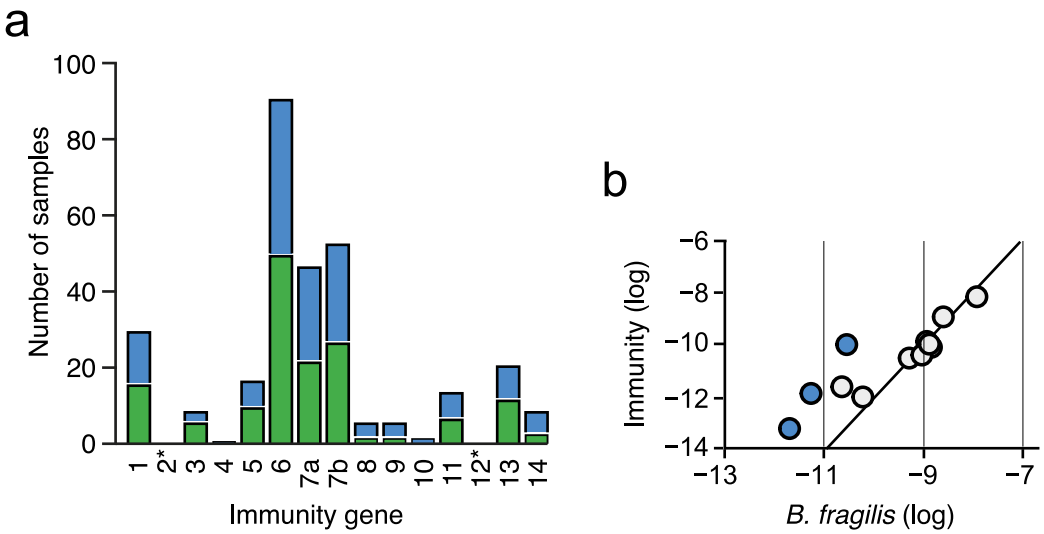
Competing interests The authors declare no competing interests.

Additional information

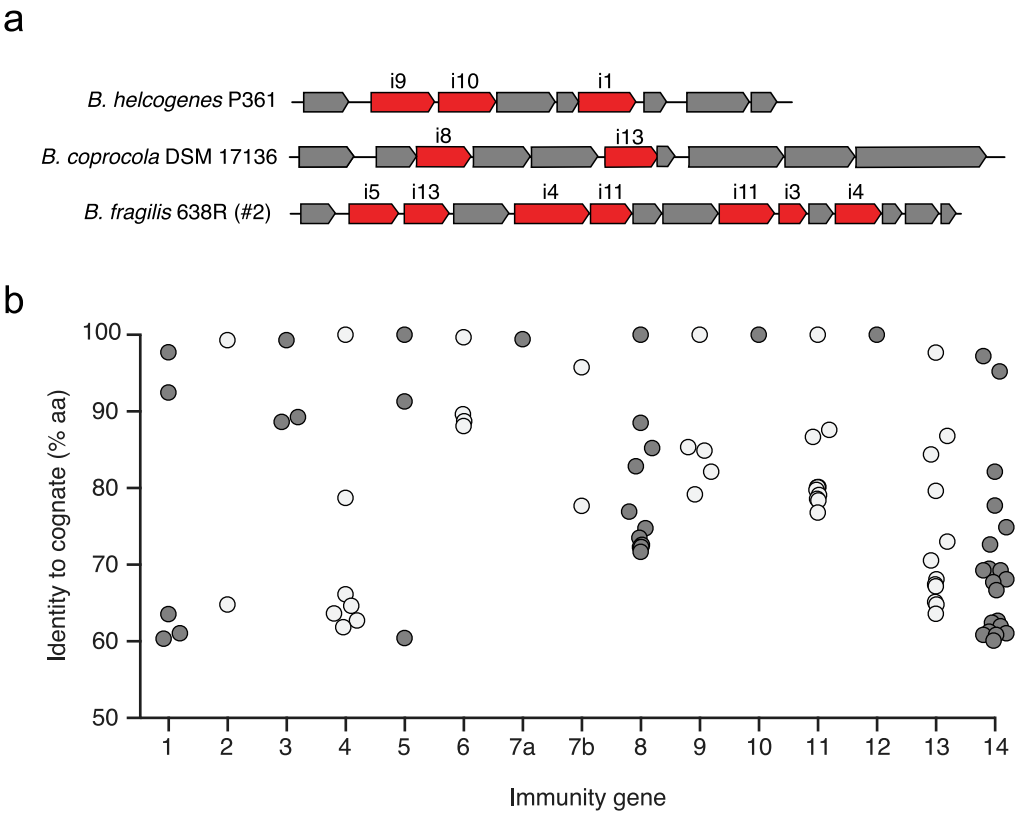
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1708-z.

Correspondence and requests for materials should be addressed to E.B. or J.D.M.

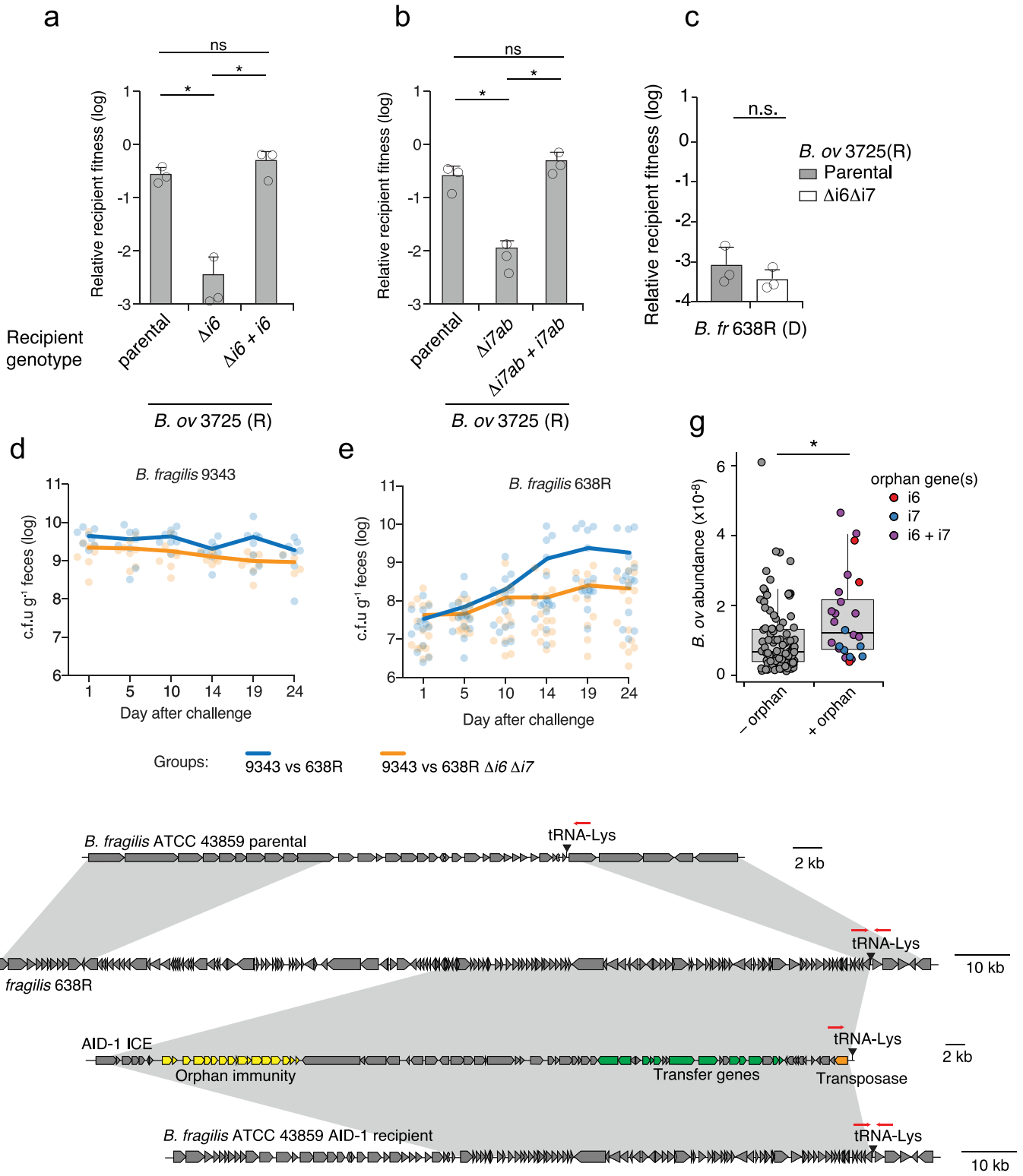
Peer review information *Nature* thanks Melanie Blokesch, Kevin Foster and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



Extended Data Fig. 1 | Prevalence of *B. fragilis*-specific orphan immunity genes in adult and infant microbiomes. a, Number of adult human gut microbiome samples in which the indicated immunity genes (1–14, GA3_i1–14 from ref. [8]) can be detected at an 80% nucleotide identity threshold and an abundance more than tenfold that of *B. fragilis* marker genes. Bars coloured as in Fig. 1a, and asterisks indicate immunity genes without orphan representation. **b**, Comparison of abundance of *B. fragilis*-specific T6SS immunity genes with *B. fragilis* species-specific marker genes in infant microbiome samples[16] (Supplementary Table 4). Abundances are calculated as in Fig. 1a. Samples in which immunity gene abundance exceeds that of *Bacteroides* by over tenfold (blue) are highlighted.

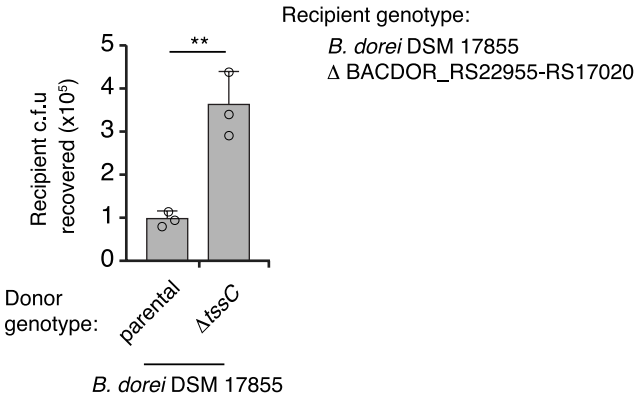


Extended Data Fig. 2 | Diversity and genomic context of orphan immunity genes in human gut microbiomes and diverse *Bacteroides* species. **a**, Representative AID-1 gene clusters containing homologues of the indicated *B. fragilis* T6S immunity genes from the indicated reference genomes. **b**, Data points indicate the amino acid identity of unique genes homologous to indicated *B. fragilis*-specific T6SS cognate immunity genes identified through BLAST analysis of the IGC[29] ($n = 88$ genes, maximum $E = 1 \times 10^{-40}$; minimum percentage identity, 60%).

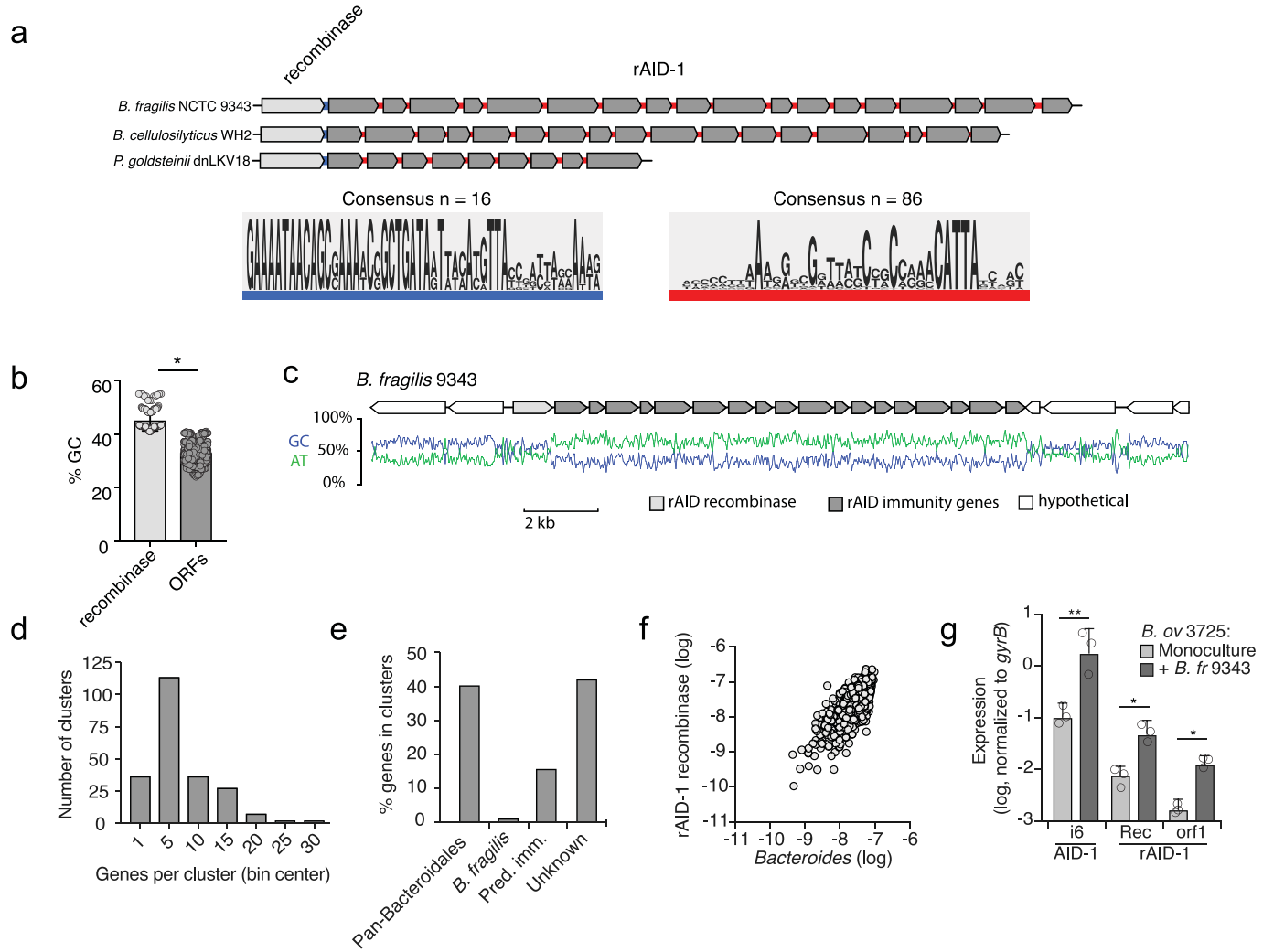


Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Orphan immunity genes specifically enhance the fitness of *Bacteroides* strains in vitro and in vivo. a, **b**, T6SS-dependent competitiveness of parental strains of *B. ovatus* 3725 and the indicated mutant and complemented derivatives during in vitro growth competition experiments with *B. fragilis* 9343. Relative recipient fitness was determined by calculating the ratio of final to initial c.f.u. and normalizing to the corresponding experiment with *B. fragilis* 9343 lacking *tssC* (T6S-inactive). Data are mean ± s.d. of three independent biological replicates. **P* < 0.01, unpaired two-tailed *t*-test. **c**, T6SS-dependent competitiveness of a parental strain of *B. ovatus* 3725 or a strain bearing in-frame deletions of indicated orphan immunity genes, during in vitro growth competition experiments with an orthogonal effector-bearing *B. fragilis* 638R parental strain or a derivative strain lacking *tssC* (T6S-inactive). Relative recipient fitness and statistics were calculated as in **a** and **b**. *n* = 3 independent biological replicates. **d**, **e**, Recovery of *B. fragilis* 9343 (**d**) or 638R (**e**) and the indicated orphan immunity mutant derivative from pairwise competitions of the strains in germ-free mice. Lines indicate the mean at each time point (*n* = 8 mice per group for each of two independent experiments). Alternating time points of these data are included in ratio form in Fig. 3c. **f**, Schematic depicting genomic loci for the *B. fragilis* ATCC 43859 parental strain, the *B. fragilis* 638R AID-1 donor strain, the AID-1 system, and the ATCC 43859 AID-1 recipient. Grey shading indicates homology; red arrows indicate the position of PCR primers used to infer insertion of the AID-1 element at the $tRNA^{Lys}$ insertion site. **g**, 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* (+). Abundances are calculated as in Fig. 1a. **P* < 0.001, Wilcoxon rank-sum test. *n* = 128 non-orphan samples, *n* = 24 samples containing orphan immunity. For box plots, the middle line denotes the median; the box denotes the interquartile range (IQR); and the whiskers denote 1.5× the IQR.



Extended Data Fig. 4 | The GA2 system of Bacteroidales mediates interbacterial antagonism. Recovery of *Bacteroides dorei* DSM 17855 cells lacking GA2_e14-i14 (BACDOR_RS22955-17020) from two-strain in vitro growth competition experiments with the indicated donor strains. $n = 3$ technical replicates representative of three biological replicates. $^{**}P < 0.01$, unpaired two-tailed *t*-test.



Extended Data Fig. 5 | rAID-1 systems include conserved and repetitive intergenic sequences and bear hallmarks of horizontal gene transfer. a, Left, motif enrichment analysis from the intergenic sequences immediately 3′ of the recombinase stop codon to the start codon of the first downstream open reading frame within 16 randomly selected rAID-1 gene clusters. This region is highlighted in blue in three representative rAID-1 systems shown above. Right, motif enrichment analysis from all 86 intergenic sequences between the ORFs of six rAID-1 clusters (*B. fragilis* NCTC 9343, *B. cellulosilyticus* WH2, *B. ovatus* 3725, *Paraprevotella clara* YIT 11840, *Parabacteroides goldsteinii* dnLKV18, and *Parabacteroides gordonii* MS-1)[44]. This region is highlighted in red in three representative rAID-1 systems shown above. **b**, Average G + C nucleotide content of rAID-1-associated recombinase versus rAID-1 predicted ORFs ($n = 226$). ***$P < 0.0001$, unpaired two-tailed t-test. **c**, Schematic depicting the G + C and A + T nucleotide content across a representative rAID-1 system from *B. fragilis* 9343. **d**, Frequency distribution of gene number in rAID-1 clusters ($n = 1,247$ genes in 226 clusters). Bin width is five genes. **e**, Composition of genes in rAID-1 clusters ($n = 226$ clusters) as determined by profile HMM scans and BLAST analysis against a curated database of Bacteroidales T6SS immunity genes[2,8]. **f**, Comparison of the total abundances of rAID-1-associated predicted recombinases and the *Bacteroides* genus in adult microbiome samples derived from the HMP and MetaHIT studies (Supplementary Table 8). Abundance values are calculated as in Fig. 1; genus abundance corresponds to the sum of all *Bacteroides* spp. (calculated individually as the average of species-specific marker gene abundances). **g**, Results of qRT–PCR analyses for the indicated *B. ovatus* 3725 genes belonging to AID-1 (i6, M088_1971) or AID-1 clusters (Rec, recombinase, M088_1401; *orf1*, M088_1400) under conditions of growth in mono- or co-culture with *B. fragilis* 9343 for 2 h. Data are mean ± s.d. of three independent biological replicates. *$P < 0.05$, **$P < 0.01$, Wilcoxon two-tailed sign-rank test.

nature research

Corresponding author(s): Mougous

Last updated by author(s): YYYY-MM-DD

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	*Provide a description of all commercial, open source and custom code used to collect the data in this study, specifying the version used OR state that no software was used.*
Data analysis	The following publicly-available software packages were used to analyze data in this manuscript: MetaPhlAn2.0, bowtie2, samtools, BLAST, SoapDeNovo2, GeneMarkS, Unicycler, HMMERv3.2.1, MEME Suite 5.0.2, GraphPad Prism 7.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All materials generated as part of this manuscript, including strains and plasmids, will be made readily available upon request following publication.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to determine sample size for in vitro experiments and experiments involving gnotobiotic mice. For mouse experiments, we used reasonable numbers considering limitations of housing and maintenance under gnotobiotic conditions.
Data exclusions	No data were excluded from reporting
Replication	All experiments reported in this study were reproducibly replicated.
Randomization	For gnotobiotic experiments, mice from multiple litters were randomized into treatment groups when possible.
Blinding	Investigators were not blinded to group allocation during gnotobiotic experiments

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	All mice used in this study were adult (8-12 weeks of age) Swiss-Webster females.
Wild animals	*Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*
Field-collected samples	*For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*
Ethics oversight	The University of Washington Office of Animal Welfare IACUC approved the protocol governing the use of gnotobiotic mice in this study.

Note that full information on the approval of the study protocol must also be provided in the manuscript.