Insertion sequence driven evolution of the microbiota revealed by a novel metagenomics pipeline

Joshua M. Kirsch[1], Andrew J. Hryckowian[2, 3], and Breck A. Duerkop[1]*

[1]Department of Immunology and Microbiology, University of Colorado - Anschutz Medical Campus, School of Medicine, Aurora, Colorado, USA

[2]Department of Medicine, Division of Gastroenterology and Hepatology, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin, USA

[3]Department of Medical Microbiology & Immunology, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin, USA

*Correspondence: Breck A. Duerkop breck.duerkop@cuanschutz.edu

Abstract

Insertion sequence (IS) elements are mobile genetic elements in bacterial genomes that support adaptation. We developed a database of IS elements coupled to a computational pipeline that identifies IS element insertions in the microbiota. We discovered that diverse IS elements insert into the genomes of intestinal bacteria regardless of human host lifestyle. These insertions targeted bacterial accessory genes that aid in their adaptation to unique environmental conditions. Using IS element expansion in *Bacteroides*, we show that IS element activity leads to insertion “hot spots” in accessory genes. We show that IS element insertions are stable and can be transferred between humans. Extreme environmental perturbations force IS element insertions to fall out of the microbiota and many fail to rebound following homeostasis. Our work shows that IS elements drive bacterial genome diversification within the microbiota and establishes a framework for understanding how strain level variation within the microbiota impacts human health.

Introduction

Bacteria in the intestinal microbiota are under constant evolutionary pressure to adapt. One way that bacteria can rapidly evolve their genomes is through the activity of mobile genetic elements, including bacteriophages, plasmids, DNA inversions, and transposable elements (1-3). Genetic rearrangements have a strong impact on bacterial fitness, influencing diverse phenotypes including virulence, antibiotic resistance, interbacterial competition, and secondary metabolism. For example, pathogenicity islands can distribute virulence traits to previously non-pathogenic bacteria and DNA inversions can alter the expression of antibiotic resistance genes in the microbiota (4, 5). Additionally, novel genetic traits such as metabolism and virulence genes are acquired by intestinal bacteria through plasmid exchange, bacteriophage lysogeny, and conjugative transposition (6, 7). Thus, mobile elements actively drive bacterial adaptation and evolution in the intestine.

Insertion sequence (IS) elements are small (~1 kB) simple transposons that are common to bacterial genomes (8). IS elements are capable of mobilizing themselves by the action of an encoded transposase that recognizes sequence-specific inverted repeats at the boundaries of the element. IS element transposases are diverse (9), and include enzymes with a DD(E/D) motif that support “copy-and-paste” and “cut-and-paste” mechanisms or those with HUH nuclease chemistry which transpose as ssDNA molecules at the replication fork (10-12). Gene expression and genome fidelity can be modulated by IS element activity in a variety of ways, including IS element insertions into protein coding regions inactivating genes, insertions into intergenic regions forming strong hybrid promoters that increase gene expression, or recombination with other IS element sequences resulting in large deletions (13-16). Additionally, IS elements can rapidly expand in bacterial genomes when commensal bacteria transition to become pathogens (15, 17, 18). Thus, IS elements have a profound impact on the evolution of bacterial genomes and the physiological traits of bacteria (19).

Despite a deep knowledge of fundamental IS element biology from diverse bacteria, we have a rather rudimentary understanding of how these elements function in polymicrobial communities. We believe this knowledge is critical for deciphering how bacterial communities evolve ecosystems such as the human microbiota. A barrier to studying IS elements in such complex environments results from imperfect methodologies for measuring *in situ* IS element dynamics. This stems from the poor recovery of multi-copy genes with repetitive sequences by short-read assemblers, leading to fragmented assemblies where IS elements are either absent or become break-points between contigs (20). Therefore, assembly-level analyses of IS elements from metagenomic datasets are often underpowered. Despite these limitations, previous culture-based studies show that IS elements are active in the human intestine. Longitudinal sampling of *Bacteroides fragilis* isolates revealed the gain and loss of IS element transposases (21). Furthermore, IS elements were shown to rapidly expand in copy number in the genome of *Escherichia coli* during intestinal colonization of mice, driving increased virulence

in a mouse model of Chron's disease (22, 23). Recently, longitudinal intestinal metagenomic samples from a single individual were sequenced using long DNA fragment partitioning, producing higher quality assemblies compared to short-read methodologies. This analysis revealed that *Bacteroides caccae* gained and lost numerous IS insertions over the course of sampling (20). Finally, we previously utilized a targeted IS element sequencing approach to identify novel insertions of the IS element IS256 in intestinal *Enterococcus faecium* populations isolated from an individual undergoing multiple antibiotic treatment (19). Specific antibiotics treatment were associated with increased IS256 abundances and enrichment of IS256 insertions in genes related to antibiotic resistance and virulence, suggesting that antibiotic treatment drives IS-mediated pathoadaptation in the human intestine.

In this work, we developed an open-source IS element database that greatly expands the diversity of IS elements compared to current databases. We built a computational pipeline that utilizes this database to find IS element insertions in a variety of public metagenomic datasets. Our analysis reveals widespread IS element insertions and expansions within the human microbiota. We show that IS element activity is transferable between individuals and are stable over extended timescales. Distinct families of IS elements are favored for insertional activity within the microbiota and these IS elements preferentially insert into specific classes of genes. These genes are specific to distinct microbial taxa with an overrepresentation within the *Bacteroidia* and *Clostridia* and are reproducible across discrete metagenomic datasets. Gene classes targeted by IS elements are primarily metabolic, cell surface, and mobile genetic element accessory genes. Using an *in vitro* ISOSDB412 (IS4351) expansion model, we confirm that identical and related accessory genes are preferentially targeted within both *Bacteroides thetaiotaomicron* and *Bacteroides fragilis*. Finally, we show that the stability of IS element insertions is lost following antibiotic perturbation and diet intervention, and following microbiota recovery after such alterations, new IS element abundances and insertion site locations arise. Together, this work establishes a framework for studying IS elements within the context of the

microbiota and is the first step toward understanding how these mobile elements contribute to the function of the microbiota during human health.

Results

ISOSDB: a comprehensive open-source IS element database

To help facilitate the study of IS elements in bacterial genomes and metagenomic datasets, we sought to build an updated and exhaustive IS element database that can serve as an open-source tool for the scientific community. Prior to our development of this database and the analysis workflow, the main repository for IS elements was the ISFinder database (24). While ISFinder has served as a gold-standard for the systematic naming of IS elements and their identification in single isolate genomes (25), it is underpowered for high-throughput genomics. Namely, ISFinder lacks integration into downstream tools that can facilitate in-depth analyses of IS elements, has accessibility issues in that the entirety of the database cannot be manually downloaded for personal use, and it relies solely on manual curation and submission of new IS elements, decreasing the speed at which new IS elements can be disseminated to the scientific community. These are serious drawbacks in the current genomic age, where the identification of IS elements from complex datasets such as metagenomes would greatly improve our understanding of their function in affecting complex bacterial traits from diverse environments.

To address the need for an unrestricted open-access IS element database to support genomic research, we built the Insertion Sequence Open-Source Database (ISOSDB), which consists of IS elements identified from 39,878 complete bacterial genomes and 4,497 metagenome assembled genomes (MAGs) (26).These IS elements were identified using OASIS, a rigorously tested IS identification tool that allows for the high-throughput analysis of multiple genomes and contigs (27). IS elements were considered valid and included in ISOSDB if there were at least two copies of the IS element in a single genome and it was flanked by

terminal inverted repeats (IRs). Redundant IS elements were deduplicated at 95% nucleotide identity. The resulting set of IS elements totaled 22,804 distinct IS elements, an almost five-fold excess to the ISFinder database (Fig. 1A). We identified transposase amino acid homologs, in addition to nucleotide homology over the whole sequence of the IS elements, in the ISFinder database (Fig. 1B). 92.6% of the transposases included in ISOSDB had homologs in ISFinder, but only 37.7% of IS elements in the ISOSDB had nucleotide homologs in ISFinder. This demonstrates that the ISOSDB has an abundance of novel IS element nucleotide sequence diversity. The ISOSDB also has a wide range of transposases representing multiple IS families (Fig. 1C). In summary, the ISOSDB is an expansive freely available database of IS elements that contains substantially improved IS element diversity compared to current IS element databases.

Development of an IS element detection pipeline that uses ISOSDB

IS elements are known to drive genomic diversity in almost every bacterial species (9, 28, 29). Considering the importance of IS elements to bacterial genome evolution, a systematic method to identify novel IS insertions in complex microbial communities has not been previously developed. This has precluded the study of IS element biology in the microbiota at the population scale. To obtain a deeper understanding of IS element biology in the microbiota, we developed the pseudoR pipeline that utilizes ISOSDB to identify IS element insertions in previously assembled genomic sequences. pseudoR is named for its ability to find IS element inactivated pseudogenes and is implemented in the R programming language. This pipeline is built for fragmented, incomplete assemblies such as fecal metagenomic assemblies, and has ample multi-threading support to speed up analyses of deeply sequenced samples. This approach was inspired by previous tools developed for the analysis of transposon insertions from eukaryotic (30) and bacterial genomes (31, 32). These tools either require *a priori* knowledge of the target site duplications which is often not consistent for IS element insertions

(33), use read pairing to infer IS element insertions which limits the data available to verify IS insertions, or require a matched reference assembly.

The backbone of the pseudoR pipeline is a split read approach. First, reads are mapped against assembled contigs and unmapped reads are binned as separate read files. These unmapped reads are aligned against a database of IS element terminal ends (150 bp on either end of an IS element) compiled from ISOSDB (Fig. S1). If an unmapped read has an IS element termini, the termini is trimmed and the remaining read is remapped against the assembled contigs. To ensure that insertions are genuine, a minimum depth of four reads is required and these reads must include at least one left and one right termini.

To compare between samples and multiple disparate assemblies, we built two modes (multi and single) into pseudoR. Multi mode takes each sample and maps its reads against its own assembly. These reads are then subsequently mapped against a deduplicated gene database built from all assemblies in the dataset. Single mode was developed to better analyze time series datasets, where a single assembly can be used for multiple samples (such as one from the start of the time series) which is key for analyzing complex assemblies from multiple individuals and timepoints.

IS element insertions are widespread in the healthy human intestinal microbiota

We first utilized the pseudoR pipeline to identify IS element insertions in healthy human fecal metagenomes (Fig. 2). To gain a comprehensive view of IS insertions in this context, we analyzed data from three geographically disparate studies: a survey of healthy Italian adults ranging in age from 30-105 years old (ITA) (34), healthy individuals from a Japanese colorectal cancer study (JPN) (35), and healthy rural Madagascan adults (MDG) (36). Each study showed evidence of widespread IS insertion heterogeneity (Fig. 2A-B). Some individuals had few detectable IS element insertions while others had over 100 unique IS element insertions (Fig. 2B). We used relative IS depth (calculated by dividing IS depth by the sum of the IS depth and

the depth of the reads that map to the insertion site without the insertion present) as a metric to report the abundance of IS alleles. Most insertions had a relative IS depth above 10% and below 100%, demonstrating that the inserted allele exists in equilibrium with the wild type allele (Fig. 2A). To determine the extent of IS element insertions within the microbiota, we determined the bacterial taxa that underwent extensive IS element insertional activity (Fig. 2C, Fig. S2A). IS element insertions were abundant in the *Bacteroidia* and *Clostridia* and insertions in these classes were found in almost every individual. These two classes make up the majority of the healthy intestinal microbiota (37) and the genera *Bacteroides*, *Phocaeicola*, *Ruminococcus*, and *Blautia* accounted for the majority of IS element insertion diversity (Fig. S2A). While *Bacteroidia* IS element insertions were present across all studies, the IS elements underlying these insertions varied between studies (Fig. 2D). Unique IS element types for the *Bacteroidia* included ISOSDB412 (IS30 family) insertions for ITA individuals, ISOSDB18121 (IS1380 family) insertions for JPN individuals, and ISOSDB33 (IS1182 family) insertions for MDG individuals (Fig. 2D). Insertions in the *Bacteroidia* from JPN and ITA individuals were formed from shared IS elements, including ISOSDB4584 (ISL3 family), ISOSDB634 (IS982 family), and ISOSDB445 (IS66 family), whereas the IS element diversity in the *Bacteroidia* from MDG individuals was mostly unique consisting of ISOSDB250 (IS66 family), ISOSDB426 (IS630 family), and ISOSDB45 (IS256 family) insertions. This suggests that regional differences in genomic makeup of the *Bacteroidia* contributes to differences in IS element content or activity that drives insertion events.

IS element insertion types among the *Clostridia* were found across all three geographic locations, with many more IS elements shared between JPN and ITA individuals (Fig. 2D). Interestingly, the Clostridia shared three IS family types (IS982, IS66, and IS30) with the *Bacteroidia*, yet the specific IS family members active were assigned to different ISOSBD groups indicating they are unique IS elements (Fig 2D). We conclude from this data that although the IS element sequence space of these two classes of bacteria are unique, they

share IS element features at the IS family level. This suggests that certain IS element families may be more promiscuous across diverse intestinal bacteria.

MDG individuals, despite having similar numbers of IS element insertions among their bacterial communities (Fig. 2B), had relatively low abundances of IS elements from *Bacteroides*, *Phocaeicola, Blautia,* and *Bifidobacterium* and much higher abundance of IS elements associated with pathobionts including *Escherichia* and *Prevotella* species (Fig. S2A) (38-41). This is supported by the presence of ISOSDB45 (IS256 family) insertions that are associated with bacteria of pathogenic potential (Fig. 2D) (19, 42-44).

IS elements can insert into coding sequences or intergenic regions of the genome which can inactivate genes or influence the expression of adjacent genes, respectively (28). To measure IS element insertions in intragenic versus intergenic sites we analyzed the precise positions of all IS elements identified across ITA, JPN, and MDG individuals by comparing the number of intragenic insertions to intergenic insertions on a per individual basis. We found that specific IS elements had preferences for intragenic or intergenic insertions (Fig. 2E). ISOSDB1159 (IS3 family), ISOSDB805 (IS3 family), ISOSDB178 (ISAS1 family), and ISOSDB237 (IS4 family) preferentially inserted into intergenic loci, while ISOSDB18121 (IS1380 family), ISOSDB445 (IS66 family), and ISOSDB412 (IS30 family) inserted more readily into intragenic loci. These data indicate that some IS elements are suited to diversify genomes through mutation whereas others may preferentially insert adjacent to coding sequences to drive phenotypic diversity by altering gene expression.

IS elements frequently insert into accessory genes important for bacterial adaptation

Having identified numerous IS element insertions within predicted coding sequences, we analyzed ORFs carrying IS element insertions, from here on called iORFs, in the ITA, JPN, and MDG individuals (Fig. 3). A deduplicated gene database of all predicted ORFs from the metagenomes from each study was assessed for iORF's. iORF's with insertions from only a

single individual were about 10-fold more abundant as iORFs with insertions from multiple individuals (Fig. 3A). This demonstrates that IS element insertions per individual have a high degree of inter-individual insertion site variation. We next performed functional annotation of the iORF's and found five broad functional categories that were shared between individuals and encompassed 20.7% of iORFs, on average (Fig. 3B). An abundance of iORFs were annotated as *susC*, *susD*, or *tonB* receptor genes in ITA and JPN individuals. These genes encode high-affinity substrate-uptake receptors for carbohydrates and cofactors (45, 46). *susC-D* homologs are frequently involved in the uptake of polysaccharides used for metabolism, with *Bacteroidia* species containing many different *susC-D* homologs (45). Mobilome genes, including transposases, integrases, prophage elements, and mobile element defense genes, such as restriction modification genes, contained insertions in almost every individual across all studies. IS element insertions in transposase and integrase genes suggests mobile element competition within bacterial genomes. Genes involved in exopolysaccharide biogenesis, such as LPS modifying enzymes and glycosyltransferases involved in cell wall synthesis were consistent IS element insertion targets in all three studies. Cell wall and membrane modifications are crucial for *in vivo* survival to avoid host immunity and during competition with other bacteria (47). Finally, antibiotic resistance genes frequently contained IS element insertions, including a high prevalence of insertions in *tetQ* and *ermF* in JPN and ITA individuals. This indicates that the carriage of antibiotic resistance genes in the microbiota is metabolically unfavorable and IS elements may modulate their activity in the absence of antibiotic pressure. Insertion abundances were enriched in these functional categories when comparing insertion abundances within all predicted protein coding sequences (Fig. 3C, Fig. S2B). Together, these data show that accessory metabolic genes and genes involved in antibiotic resistance and genomic plasticity are common IS insertion targets and indicates that the intestinal microbiota modulates the function of these genes to adapt to their environment.

We next asked if certain iORFs were shared between individuals from the three studies. Although infrequent, a few iORFs were shared. 14% and 37% of ITA and JPN individuals, respectively, had IS insertions in the gene *ermF*, a macrolide resistance gene, and 3.2% and 9.8% of ITA and JPN individuals, respectively, had IS insertions in *KAP*, a P-loop NTPase predicted to be involved in phage defense (Fig. 3D) (48, 49). 7.9% and 6.1% ITA and MDG individuals, respectively, shared IS insertions in *shufflon*, a gene encoding an invertase that regulates pilin phase variation (Fig. 3D) (50). We found that iORF functional classes were conserved across diverse intestinal bacteria (Fig. 3E) yet the IS element types responsible for these insertions differed among the various classes of bacteria (Fig. 3F). This demonstrates that although IS elements target similar genes within disparate bacteria, the types of IS elements that promote these diversification events are specific to certain classes of bacteria.

Multiple iORF's were classified as transposases (Fig. 3B) and examples of IS elements inserting in other IS elements has been previously reported (51). In order to understand the dynamics of IS insertions in IS elements, we classified both the transposase iORF and the IS element forming the insertion by IS family (Fig. 3G). We found multiple examples of both self-targeting (an IS element inserting into a closely-related IS element) and orthologous-targeting (an IS element inserting into an unrelated IS element). We found that some IS families such as ISL3, IS4, IS110, and IS256 experienced promiscuous insertion events. Interestingly, IS elements from some of these families create composite transposons capable of mobilizing cargo genes including metabolic and antibiotic resistance genes (52-55). This observation is consistent with an enrichment of iORFs belonging to both metabolic and antibiotic resistance accessory genes and suggests IS element self- and orthologous-targeting could be a mechanism to control IS element mobilization of genetic traits when these events are unfavorable to the bacterial host.

IS elements provide mutational diversity to Bacteroides species

Recently a genome wide mutagenesis study of the *Bacteroides thetaiotaomicron* chromosome identified genes that support adaptation to environmental pressures, including antibiotics, bile acids, and carbon sources (56). We used this dataset to evaluate whether we could identify similar fitness determinants as iORFs within the natural *Bacteroides* population within the intestine. We searched our iORF database for homologs to these fitness-associated ORFs from *B. thetaiotaomicron* and found multiple closely related iORFs that likely influence *Bacteroides* fitness both positively and negatively (Fig. S3A). Examples include iORFs that provide fitness advantages during growth in the presence of the antibiotics doxycycline and vancomycin and during hyaluronic acid and glucosamine consumption (Fig. S3A). Interestingly we found iORFs associated with bacterial fitness during exposure to the antipsychotic medication chlorpromazine (Fig S3A). Antipsychotic compounds are connected to alterations in the intestinal microbiota and directly inhibit bacterial growth as well as potentiate antibiotic resistance (57-60). Additionally, we found multiple IS element insertions in homologs of the *B. thetaiotaomicron susC* gene BT1119 (Fig. S3B). Inactivation of this *susC* gene decreases fitness during growth on galacturonic acid, a downstream product of pectin degradation (56).

Our *in silico* analysis shows that the intestinal *Bacteroidia* experience large-scale IS element mobilization into discrete groups of genes including *susC-D/tonB* carbohydrate utilization genes as well as genes involved in exopolysaccharide synthesis and mobile genetic element interactions (Fig 2, Fig. 3). These data indicate that these genes are “hotspots” for IS element insertion. To test if the activation of an IS element leads to predictable insertion into these specific classes of genes, we transformed a multi-copy plasmid carrying an *ermF* cassette disrupted with a copy of ISOSDB412 (IS30 family member IS4351) into *B. thetaiotaomicron* VPI-5482 (*Bt*) and *Bacteroides fragilis* NCTC 9343 (*Bf*), both which lack native ISOSDB412. Following the growth of these transformants, we isolated their genomic DNA and performed IS-Seq, a technique that enriches for ISOSDB412 amplicons during Illumina DNA sequencing library construction and read mapping to identify specific ISOSDB412 insertion locations (19).

We found that both *Bt* and *Bf* strains carrying the ISOSDB412 plasmid obtained numerous ISOSDB412 insertions throughout their genomes compared to controls that lacked the ISOSDB412 plasmid construct (Fig. 4A, Fig. 4D). As predicted, ISOSDB412 insertions were enriched in gene categories reflected from our findings from insertions from the human microbiota (Fig. 4B, Fig. 4E). Specifically, for *Bt* there was an enrichment of ISOSDB412 insertions in *susC-D/tonB* and EPS biogenesis genes, whereas *Bf* acquired an abundance of insertions in *susC-D*/*tonB* genes despite a significantly lower number of I ISOSDB412 insertion events compared to *Bt* (Fig. S4A). Additionally, insertions were more frequently found in intragenic loci compared to intergenic loci, confirming our findings from the intestinal microbiota (Fig. 4C, Fig. 4F). These results demonstrate that IS elements can rapidly diversify *Bacteroidia* genomes.

Our discovery that IS insertions target bacterial accessory genes suggests that these IS insertions inactivate genes that are currently dispensable for survival in the intestine. To explore this, we measured IS insertional dynamics in a nutrient-deplete environment with readily-digestible glucose as the sole carbon source. We hypothesized that a lack of fiber and complex carbon sources, high levels of glucose, and single sources of iron and nitrogen, would incentivize higher rates of IS insertions to inactivate costly metabolic machinery needed for competition in complex nutritional environments. We measured ISOSDB412 insertion events from three *Bt* colonies after three sequential passages in either the complex media (BHIS) or minimal media (MM) (Fig. 4G, Fig. 4H, Fig. 4I). We found that colonies passaged in MM had substantially more IS4351 insertions compared to BHIS-passaged colonies. Numerous insertions were found in the open reading frame BT3642, a Na^+-dependent transporter (Fig. S4B) in MM colonies, but not in BHIS colonies. This gene has been shown to be detrimental for growth in glucose minimal media (56). Next, we measured ISOSDB412 abundance in *Bt* cells chronically infected with the *Crassvirales* DAC15 or DAC17 (Fig. 4J, 4K, 4L). *Crassvirales* are bacteriophages that selectively infect *Bacteroides* species and are highly prevalent in the

human intestinal microbiota (61). These strains shed high levels of infectious phage particles following chronic phage infection (Fig. S4C). While DAC15 and DAC17 share 99.24% nucleotide identity over 98% of their genomes, DAC17 infection is associated with an increased number of insertions genome-wide (Fig. 4J, Fig. 4K) and preferentially inserted into genes involved in EPS biogenesis (Fig. 4L). Furthermore, almost all of the infected strains carried IS insertions in the CPS3 locus (Fig. S4D), which is crucial for productive phage infection and likely leads to phage resistance (62). Together, these results show that IS element diversification of the *Bt* genome is more prevalent under nutrient-limited conditions and during infection with intestinal-resident phages. This indicates that IS element activity supports the ability of *Bt* to adapt when faced with fitness constraints.

Temporal monitoring of IS element activity shows that IS insertions are maintained over time.

Our assessment of IS element insertions in members of the intestinal microbiota revealed that IS insertions in specific genetic loci coexist with the wild type allele lacking the IS element (Fig. 2). This indicates that IS element insertions in target loci represent only a subpopulation of bacterial cells within the community and that these insertions are maintained in equilibrium with bacterial cells harboring uninserted versions of the loci. To understand the temporal longevity of IS element insertions in the microbiota, we analyzed longitudinal fecal metagenomic samples from healthy individuals (63, 64). Using a set of two individuals whose fecal samples were collected over the course of 76 and 91 weeks (63), we measured the IS element landscape of their microbiotas using pseudoR (Fig. 5A & Fig. S5). We also analyzed longitudinal samples from an additional 10 individuals from a separate study (64) (Fig. S6). For each individual, we used the assemblies from the first timepoint (indicated as week 0 in Fig. 5A) as the reference for the pseudoR pipeline which was then compared to all other time points to assess 1) the maintenance of the IS elements, and 2) the acquisition of new IS elements that

were not found at the starting timepoint. We found that most IS element insertions are maintained for periods extending to up the endpoint of the study for each individual and that some of these insertions frequently went in and out of detection (Fig. 5A, Fig. S5-6). New insertions not present at week 0 arose at almost every timepoint, with a variable number of insertions per timepoint. Some new insertions were maintained for extended periods of time while others were only detected at the initial timepoint of week 0. These results demonstrate that IS elements are both stably carried in the microbiota and are actively forming new insertions.

Next, we analyzed whether the host bacterium's abundance could account for the detection of new IS elements (Fig. S7A). Increasing levels of an IS element's host bacterium would generate higher sequencing read depth and more detection power for new IS insertions. For every insertion not present at week 0, we compared the insertion site's read depth (a marker of host abundance) to the maximum insertion site depth of all earlier timepoints (prior to the detection of the insertion). Insertions with insertion site depths equal to or less than 200% of the previous maximum insertion site depth were considered to arise from new insertional activity (Fig. S7B). We found that the majority of new insertions could be accounted for as active IS element insertions and not increased bacterial host abundance (Fig. S7C).

After week 0, multiple insertions were never detected again. To confirm that these insertions were lost from the population rather than being undetectable due to low sequencing depth, we compared the WT allele depth of insertions that were present at week 0 and absent at every other timepoint. Insertions where the WT allele depth reached at least 50% of the week 0 WT allele depth at a later timepoint were classified as "rebounders" and those that did not were classified as "non-rebounders". We found that, for every insertion that fit the above criteria, the WT allele depth rebounded to at least 50% of the week 0 WT allele depth (S7D). This suggests that these insertions were lost due to competition from the WT allele, rather than a change in read depth.

Insertions across the time series included diverse IS element families and were frequently associated with the *Clostridia* and *Bacilli* (Fig. 5B, S5-6). Additionally, we observed that the rate of new insertions per week for each individual ranged on average from around 1 to 10 new insertions and that the maintenance, loss, and gain of IS elements tracked with common bacterial accessory genes that we established to be “hot spots” for IS element insertions (Fig. 5C, Fig. 5D).

IS elements are efficiently transferred and maintained within new individuals

Having discovered that individual microbiotas have unique patterns of IS element insertions, we wanted to understand how the human host influences the maintenance, loss, and gain of IS elements within the microbiota. To test this, we used pseudoR to profile the IS element insertions during a fecal microbiota transplantation (FMT) where the microbiotas of FMT donors and recipients were longitudinally sampled before and after fecal transplantation (63). We compared the rates of IS element insertions and their maintenance in both the donors and recipients using the assembly of the donor’s transplanted microbiota as a reference. Donor-derived communities harbored both insertion site DNA with and without an IS element insertions and were stably maintained over one year in the recipients that received the FMT (Fig. 6A & Fig. S8). Insertions present prior to transplantation in recipients were lost, but a minority returned at later timepoints. Recipients maintained significantly less IS element insertions that were present at week 0 compared to the donors (Fig. 6C), but both donor and recipients had similar rates of new insertions (Fig. 6D). To understand if IS element insertional activity was similar between host and recipient, we compared the pattern of new shared insertions in donors and recipients at all timepoints post transplantation (Fig. 6B & Fig. S9). We found multiple instances of the same newly detected insertion arising in both donor and recipients at the same timepoint. These results demonstrate that complex IS element insertions can be stably maintained in new hosts

for extended periods of time and that IS element mobilization occurs at similar rates and patterns in new hosts.

Antibiotics drive widespread loss of IS insertions within the microbiota

Considering IS element insertions were found to persist for prolonged periods of time within the microbiota and were portable between individuals, we wanted to know what would happen to IS element diversity if a stable community was disrupted. To test this, we analyzed a microbiota dataset for IS element composition before, during, and after broad spectrum antibiotic treatment (65). The metagenome of the microbiota preceding antibiotic treatment was used as the reference for IS element comparison using pseudoR. To account for IS element activity within the microbiota prior to antibiotic treatment and to exclude the possibility that new insertions were caused by large shifts in rebounding bacterial abundances following antibiotic treatment, new insertions had to meet the selection criteria established where the insertion site lacking an IS element must have a sequencing depth less than or equal to 200% compared to the reference insertion site prior to antibiotic treatment. As expected, antibiotic treatment caused a substantial loss of IS element diversity, which was not fully restored after treatment (Fig. 7A & Fig. S10A). The starkest example was sample ERAS10. Extensive IS element diversity was present at week 0 in ERAS10 before antibiotic treatment and by at week 0.6 there were no detectable IS element signatures nor did any of the original IS elements return by week 26 post antibiotic therapy. In contrast, samples ERAS1 and ERAS12 both had a complete loss of pre-antibiotic treatment IS element insertions immediately following treatment, yet a considerable number of pre-antibiotic treatment IS element insertions returned at later timepoints. Across all samples, IS element insertions in *Clostridia* genomes showed a dramatic loss of original IS element signatures, similar to the pattern seen in ERAS10 (Fig. 7C). To better understand the mechanism behind the loss of week 0 insertions, the wild type allele depth of week 0 insertions that were never detected again was plotted against time (Fig. 7B & Fig. S10B). Again, lost

insertions were classified as either non-rebounders, where the insertion site DNA sequencing depth at later timepoints never reached 50% of the pre-antibiotic treatment insertion site read depth, or rebounders, where the insertion site DNA sequencing depth at later timepoints was greater than or equal to 50% of the pre-antibiotic treatment insertion site read depth. Rebounders are indicative of the wild type allele out-competing the IS inserted allele, while non-rebounders are indicative of complete loss of the host bacterium. Examples of both non-rebounding and rebounding alleles were found for every individual. Finally, we compared the rate of maintenance of pre-antibiotic treatment insertions in individuals treated with antibiotics and those who were not. Antibiotic treatment was associated with a significant decrease in the ability of the microbiota to maintain IS insertions (Fig. 7D), due to both competition from wild type alleles and loss of host bacteria.

IS element insertion into susC-D/tonB loci is associated with diet intervention.

Our analysis has identified *susC-D/tonB* genes as major targets for IS element insertions, suggesting that IS elements can modulate carbohydrate metabolism in the *Bacteroidia* (Fig. 3), and we showed in *B. thetaiotaomicron* that nutrient stress can promote these insertions events into predictable metabolic loci (Fig. 4). With these data in hand, we suspected that a radical diet intervention in humans would lead to a shift in IS insertions in metabolic-associated genes and particularly *susC-D/tonB*. To test this hypothesis, we analyzed the IS element insertions among the intestinal microbiota of obese individuals before, during, and after a rigorous, one year-long diet intervention (66). Participants in this study received a low-calorie formula-based diet for 12 weeks, followed by a 6-week period of phasing out the formula diet for solid food without changing total caloric intake, followed by a 7-week phase where caloric intake was increased while preventing weight gain (67). Similar to antibiotic treatment (Fig. 7), we found that diet change was associated with frequent loss of the starting set of IS insertions at week 0 (Fig. 8A, 8E & Fig. S11A). Loss of the target insertion site without

the IS element was assessed in addition to the loss of IS element from these sites. In contrast to antibiotic treatment, read depth at target insertion sites lacking an IS element usually rebounded, suggesting that loss of week 0 insertions is due to competition from wild type alleles rather than loss of the host bacterium (Fig. 8B & Fig. S11B). *susC-D/tonB* genes import specific carbohydrates and we suspected that these genes were the targets for IS-mediated insertion events during diet intervention, thus we reasoned that the presence of certain carbohydrates would increase the fitness burden of an insertion in that carbohydrate's cognate *susC-D/tonB* gene, promoting loss of the IS element from this location. Alternatively, low abundance of a carbohydrate would favor insertions in the carbohydrate's cognate *susC-D/tonB* genes to reduce the fitness burden of maintaining the gene's expression. We found that ancestral *susC-D/tonB* insertions were frequently lost during the course of the diet study, new insertions in *susC-D/tonB* genes arose after the diet intervention ended (Fig. 8C). In many cases, insertions arose in the *susC-D/tonB* homolog BT1119 which we identified as an insertion site of preference (Fig. 8C, 8D). Together, these results demonstrate that diet is associated with IS insertion patterns that specify alterations in insertion events in *susC-D/tonB* genes.

Discussion

IS elements are fundamental to the evolution of bacterial genomes. They likely play a key role in the assembly of polymicrobial communities and influence how bacterial members of such communities interact. Despite this, IS element diversity and function within polymicrobial communities like the intestinal microbiota are poorly defined and understudied. Here, we built a novel computational pipeline for identifying IS element insertions within complex metagenomic DNA sequences. This pipeline, which we call pseudoR, was built using a custom open-source IS element database (ISOSDB), which vastly improves nucleotide diversity compared to current state-of-the-art databases for IS element analysis. Using pseudoR, we found that IS elements contribute to high levels of bacterial genomic diversification within the microbiota, are

maintained as integrated elements at specific loci for extended periods of time, can be efficiently transferred between individuals, and experience both losses and gains within bacterial populations following perturbation (i.e., antibiotics treatment or diet alteration). Together, these results demonstrate that IS elements are important genetic elements that can dictate genotype diversity within the microbiota. The functional consequences of IS element activity within the microbiota are just beginning to be understood and should prove to have a strong impact on the microbiota and its relationship to human health.

Using previously published metagenomic datasets, we showed that the majority of open reading frames containing IS element insertions (iORFs) are unique and not shared between different individuals (Fig. 3A). This is similar to the observation that single nucleotide polymorphism (SNP) diversity in the human intestinal microbiota is more likely to be unique than shared (68). Our work expands upon this idea and suggests that each person's "fingerprint" of microbial diversity as it relates to genotype is also strongly influenced by IS element activity.

While iORFs are unique within individuals, the functional classes of genes that constitute these iORFs are consistent between individuals (Fig. 3B). iORFs were frequently annotated within five broad categories: *susC-D/tonB*, mobilome primally composed of prophage elements and transposase genes, EPS genes, antibiotic resistance genes, and genes implicated in anti-MGE systems. Genes belonging to these categories are frequently found as accessory genes and are often non-essential. Interestingly, genes regulated by DNA inversions including those regulated by phase variation fall into similar functional categories compared to those targeted by IS element insertions (69, 70). This suggests that insertions and inversions control gene expression of the microbiota to position bacteria for optimal adaptation to changing environmental conditions. Additionally, *susC-D*/*tonB* receptors are often targets of nucleotide evolution in the human intestine possibly to help gain new abilities based on substrate availability (21). In support of this, a SusC epitope was shown to active T cells to promote an anti-inflammatory IL-10 response in healthy individuals and a proinflammatory IL-17A response

in people with Chron's disease (71). This suggests that IS elements could influence the immunomodulatory capacity of the microbiota depending on whether they are inserted within genetic loci such as *susC* homologs. Carbohydrate and other metabolite availability can have a strong influence on the virulence of intestinal of pathogens (40, 72, 73). Since *susC-D/tonB* are most prevalent among the *Bacteroidia* which dominate the intestinal microbiota and participate in the utilization of carbohydrate substrates, IS element insertions in *susC-D/tonB* genes could effectively shape the nutrient landscape of the intestine promoting or subverting pathogen fitness. In support of this theory, we showed that IS elements frequently target *susC* genes homologous to BT1119 from *B. thetaiotaomicron* implicated in the transport galacturonic acid. Galacturonic acid inhibits the activation of the type III secretion system of enterohemorrhagic *E. coli* and reduces its virulence (74), thus the inactivation of BT1119 by IS element insertion may result in regulated control of intestinal galacturonic acid concentrations that can provide protection against invading pathogenic *E. coli*.

Often, the microbiome of urban individuals is less-well suited to degrade diverse polysaccharides compared to rural individuals (75). Here, we found that urban individuals (those in the ITA and JPN cohorts) have more extensive modulation of the microbiome through higher rates of IS insertions in *susC-D/tonB* genes, compared to individuals living in rural settings (MDG cohort) (Fig. 3E). This could partly be explained by the higher ITA and JPN individuals have almost exclusively *Bacteroidia* IS insertions in *Bacteroides* and *Phocaeicola* species, while MDG individuals have *Bacteroidia* insertions in *Prevotella* species. However, we expect that the varied diet of MDG individuals, which heavily includes wild meat and plants (76), partially selects for a diverse set of functioning *susC-D/tonB* receptors. In support of this finding, diet intervention modulated IS insertions in *susC-D/tonB* genes (Fig. 8). Further work to better characterize the IS insertional landscape in individuals with diverse lifestyles, such as the Hadza hunter gatherers (77), can better demonstrate how IS elements impact nutrient acquisition.

Frequent IS element inactivation of diverse MGE systems has been observed and was hypothesized to help protect host bacteria from the toxic effects of these systems or to allow incoming MGEs to subvert these systems (33). We found that IS element insertions into other MGEs was frequent within the microbiota and that these insertions often co-exist with the insertion site lacking the IS element. This suggests that IS elements exist in an equilibrium between the inserted and non-inserted state at the population scale to maintain effective host defense against MGEs and when dispensable, relax the cost of maintaining these systems through inactivation.

In addition to metabolic and MGE system targeting by IS elements, another common iORF was *ermF*, a macrolide resistance gene that acts by methylating the 23S rRNA subunit to inhibit protein translation (78). We previously found that gentamicin resistance operons are frequently interrupted by IS elements during enterococcal infection (19). We presume that in the absence of antibiotic pressure, bacteria can dampen their effort toward maintaining high-level antibiotic resistance at the population level to favor their fitness. However, since these are heterogenous populations that co-exist between antibiotic resistance genes carrying or lacking IS element insertions, we hypothesize that IS element insertions in *ermF* and other antibiotic resistance genes serves as a mechanism to maintain a low level of population-wide antibiotic resistance, while reducing the burden of expressing the resistance gene. Such a “hedge betting” strategy would be advantageous upon antibiotic exposure where the population would not require genetic mutation to overcome antibiotic pressure but instead regain antibiotic resistance thorough selection of variants within the population.

Prior to this study, it was not known whether IS element insertions could be stably maintained in bacterial genomes within host associated microbiotas. Having this knowledge provides new insights into the genomic composition of the microbiota and their capacity to harbor stable subpopulations of genetic variants that are prime for selection and expansion under changing environmental conditions. This could account for the rapid adaptability of the

microbiota and its resiliency to recover from perturbation. We found that specific IS element insertions were detectable for just under two years, demonstrating that IS insertions within specific genetic loci are often not sanitized from the population. Additionally, new previously undetected IS elements were detectable and maintained for extended periods of time. Such insertions were likely due to new transpositions, rather than bacterial host immigration within the microbiota or changes in preexisting host bacterial abundances. These findings are analogous to the discovery of SNP conservation within the microbiota (79). Together, our results show that in addition to active and widespread transposition, IS element insertions in the microbiota are maintained for prolonged periods of time which may support rapid responses to bacterial community assembly under selective pressure and/or fluctuations in environmental resources.

In addition to longevity, we also established that IS elements can be stably transferred between individuals. Donor-derived bacterial communities have been shown to persist in recipients following FMT (80-82). Bacterial diversity frequently rebounds after antibiotic treatment (65, 83), but how bacterial communities assemble at the strain level is more varied and includes widespread loss of mutations and enrichment of others (84). We found that IS insertions were efficiently transplanted between individuals (Fig. 6), but were readily lost after antibiotic treatment and diet intervention (Fig. 7-8). We suspect that IS element insertions impart only a minor impact on bacterial fitness under steady state conditions when selective pressure is low, but become burdensome during strong selective pressure when environmental conditions are dramatically altered. Furthermore, the dysbiosis that is caused by antibiotic treatment and diet fluctuations can eventually resolve to the previous taxonomic composition (65, 66, 85, 86) . We show that in contrast to higher level taxonomic restoration of the microbiota following perturbation, IS element insertions are irreversibly altered following strong selective pressure (Fig. 7-8). This indicates that many IS element insertions within the population come with fitness tradeoffs that can be detrimental when the appropriate selection is applied. It would also be fair to speculate that some IS element insertions would be beneficial

and become dominate and fixed within the population depending on the type of environmental pressure applied. Our results suggest that IS elements can be viewed as strain-level features that are more variable than taxonomic-level diversity following perturbation.

Antibiotic exposure enriches IS insertions in *E. faecium* during a chronic human infection and antibiotics can stimulate IS element activity in *Staphylococcus aureus* (19, 87). In each of these scenarios this activity was linked to the IS element IS256, which is strongly associated with pathogenic and hospital-associated bacteria (18). In contrast, we found that during antibiotic treatment the intestinal microbiota undergoes extreme reductions in IS element diversity among the commensal bacterial population. This suggests that IS element activation by antibiotic pressure could be dependent on the specific family of IS element or certain pathogenic bacteria have developed IS elements as a tool to mutagenize their genomes in response to antibiotic stress. The latter strategy could help support pathogen fitness in a context dependent manner (i.e. infection), for which there is little to no benefit to commensal bacteria during intestinal colonization.

In summary, the development of a contemporary and stringently curated IS element database, combined with a streamlined computational approach to identify IS elements from metagenomic data, has revealed novel insights into how IS elements shape the genomes of the microbiota and sets the stage for exploring how specific IS element variants in the microbiota influence human health. However, a few limitations to the tools and study exist. First, delineation of IS element insertion sties between closely related strains is imperfect, thus insertions between very closely related strains may group together and fail to resolve. Work from the intestinal microbiota has found that most species are represented by one dominant strain and that these strains are stable over long periods (88), suggesting that the insertion sites that we measured are from subpopulations originating from single strains. Second, we do not understand how IS element insertions influences the fitness of the microbiota. We can begin to overcome this problem by assembling and assessing Tn-Seq libraries of diverse intestinal

bacteria under conditions relevant to the intestinal environment and in animal colonization experiments. Coupling such studies with pseudoR analyses of microbiota datasets would begin to help determine what IS element insertion sites are beneficial or detrimental to commensal bacterial genome evolution and would aid in the identification of IS element families that are important for such evolutionary events in the intestine. Third, the pseudoR pipeline relies on assemblies containing some level of wild type allele being measured in comparison to an IS element insertion at that same allele. This limits the detection of IS elements that assemble as part of the reference contig. Conceivably, an inverse method to find wild type alleles in IS-allele assemblies can be built and be implemented into a future release of pseudoR. Fourth, the single reference mode of pseudoR excludes assemblies from later timepoints to prevent over-assembly and focus the analysis on the initial timepoint. This hinders our understanding of how IS elements are impacting the microbiota not present at the initial timepoints.

Materials and Methods

Creation of the ISOSDB

All genomes classified as “complete” in NCBI Assembly were downloaded on 05/08/2023 using bit v1.8.57 (89). IS elements were identified using OASIS (27). This program uses annotated transposases as points for sequence extension and comparison to find complete IS elements. Identified IS elements were filtered using the following criteria: 1) multiple (minimum of 2 copies) complete copies of the element must be present in at least one genome and 2) the elements must contain identified inverted repeat sequences. To supplement these genomes, we also identified IS elements in hybrid-assembled MAGs from a large human intestinal microbiota study (26). These metagenomes were annotated with Prokka v1.14.6 (90), processed as described above, and then combined with the IS elements identified from the genomes from NCBI Assembly. Redundant IS elements in the final database were deduplicated using CD-HIT-EST (91) with a 95% sequence cutoff and a word length of 9. IS element

homologs in ISFinder were found using either blastn (complete elements) with a minimum e-value of 0.000001 and a minimum length of alignment of 224 bp or blastp (transposases) with a minimum e-value of 0.000001 (92). Phylogenetic analysis of transposase genes was performed by multiple sequence alignment using MAFFT v7.520 (93). A phylogenetic tree was created using FastTree with default settings (94) and visualized using iTOL (95). The ISOSDB and pseudoR pipeline is freely available at https://github.com/joshuakirsch/pseudoR. Each IS element in this database is given a unique numeric identifier.

pseudoR pipeline

To characterize IS elements and their insertion sites in metagenomic DNA sequences we built the pseudoR pipeline. This tool relies on deeply-sequenced short read Illumina data and can be used for both metagenomic and single isolate genome data. We found a minimum of ~20 million paired-end reads is sufficient for identifying IS elements using pseudoR. Data input into the pipeline are deduplicated reads and assembled contigs. For this study, reads were downloaded from NCBI SRA (96) using fasterq-dump v2.11.0 from the studies previously mentioned (34-36, 63, 65, 66). Study accession numbers are provided in Table S1. Duplications, quality trimming, and read decontamination was performed using programs that are part of the BBTools software suite (97). PCR duplicated reads were removed with clumpify.sh with the flag “subs=0”. Adapter and quality trimming was performed with bbduk.sh and human, mouse, and phi29 contamination was removed using bbsplit.sh. Cleaned and deduplicated reads were assembled into contigs using MEGAHIT v1.2.7 using the “meta-large” preset (-k-max 127 -k-min 27 -k-step 10) (98). Contigs input into the pipeline include those equal to and greater than 1 kB. When using the multiple reference mode, assemblies were combined and ORFs were predicted using pprodigal.py (a parallelizable wrapper of prodigal (99)) and deduplicated using dedupe.sh. If using the single reference mode, only the reference assembly is annotated as described above. Nucleotide sequences of the predicted ORFs were used for

read mapping using Bowtie2 v2.4.5 (100). To identify IS insertions within the microbiota from longitudinal samples from the same individual, we built the single reference mode, where the individual's starting timepoint assembly is used to find novel IS element insertions. This approach reduces the chances that meaningful information is lost during clustering of multiple assemblies from multiple timepoints. In addition to time-series, the single reference mode can be used for evolution studies where a single population is compared between different treatments. Multiple reference mode is used for comparing the insertional patterns between multiple individuals. To identity IS elements in multiple reference mode, reads were aligned to each sample's own assembly using Bowtie2. Any mapped read, including discordantly mapped reads, were considered positive hits. Unmapped reads were collected and aligned to a database of IS termini using blastn and then filtered using a custom R script. This filtering script ensures that the IS termini is properly positioned on the read such that the remainder of the read is outside of the IS element. The IS termini are removed and the trimmed reads were then re-mapped to the sample's assembly as unpaired reads using Bowtie2. Following this, the original mapped reads and IS-termini trimmed reads are mapped against the deduplicated gene database. For single reference mode, reads from every sample are mapped to the provided reference assembly and the remainder of the steps are the same. Following read mapping completion, a custom R script is used to identify DNA insertion sites with read mapping from both the left and right terminus of the IS element at most 20 bp away from one another and this script compiles read abundances for these regions. Mosdepth v0.3.3 (101) is used to determine insertion site sequencing depth for the loci described above. Seqkit v2.2.0 (102), samtools v1.6 (103), R, ggplot2, tidyr, and MetBrewer were used for various computational tasks. Scripts to reproduce all data analysis presented here are available in the GitHub repository above.

iORFs assoated with B. thetaiotaomicron

B. thetaiotaomicron VPI-5482 coding sequences were downloaded from fit.genomics.lbl.gov. Homologs to iORFs were found using blastp with a maximum e-value of 0.00005 and a minimum identity of 75%. Hits to multiple related fitness determinants were filtered to include the representative with the strongest fitness value.

Bacteria

Bacteroides thetaiotaomicron VPI-5482, *Bacteroides thetaiotaomicron* VPI-5482 CPS3, and *Bacteroides fragilis* NCTC 9343 were cultured as previously described (104). Briefly, these strains were grown in brain-heart infusion (BHI) broth (Becton Dickinson, Franklin Lakes, NJ) supplemented with 1 g/L cysteine, 5% w/v $NaHCO_3$, and 5 mg/L hemin (BHIS), Varel-Bryant minimal media (105) supplemented with 28 mM glucose, or *Bacteroides* phage recovery medium (BPRM) (106) at 37° C in a Coy Type A vinyl anaerobic chamber in an atmosphere of 5% hydrogen, 20% carbon dioxide, and balanced nitrogen (Coy Lab Products, Grass Lake, MI). For growth on solid-media we used BHI agar supplemented with 10% defibrinated calf blood (Colorado Serum Company, Denver, CO). *Escherichia coli* S17-1 was grown in Lennox L broth (Fisher Scientific, Hampton, NH) with aeration at 37° C. Tetracycline was used at 2 µg/mL, gentamicin was used at 200 µg/mL, and ampicillin was used at 100 µg/mL.

DNA manipulation and cloning

Primers and plasmids used in this study are listed in Table S2. All plasmid constructs were made using Gibson assembly master mix (New England Biolabs, Ipswich, MA). Assembled constructs were electroporated into electrocompetent *E. coli* S17-1. These cells were prepared by washing cell pellets three times in ice-cold electroporation buffer (0.5 M sucrose, 10% glycerol). The resulting transformants were mated with *B. thetaiotaomicron* or *B. fragilis* by mixing 500 µL of stationary phase *E. coli* S17-1 and 500 µL of stationary phase *Bacteroides* on BHI blood agar overnight at 37° C under aerobic conditions. Transconjugants

were selected by scraping the bacterial lawns into BHI and plating serial dilutions on BHI blood agar supplemented with tetracycline and gentamicin, and incubating at 37°C anaerobically for 48 hours. The ISOSDB412 (IS4351) DNA sequence construct was purchased from Twist Biosciences (San Francisco, CA). pB006 (Addgene plasmid #182320), a gift from Dr. Lei Dai, was used as a shuttle vector and purchased from Addgene (Watertown, MA) (107). The *tetM* cassette was cloned from pCIE-*tetM* (108) and the *ermF* cassette was cloned from pG10K, a gift from Dr. Janina Lewis (Addgene plasmid #191377) and purchased from Addgene (109).

DNA extraction and IS-Seq

Genomic DNA was extracted from 1.5 mL cultures of *Bacteroides* strains using the Zymobionics DNA Miniprep kit (Zymo Research, Irvine, CA) and used as input for IS-Seq as described previously (19). Briefly, NGS libraries were produced using the Illumina DNA Prep Kit (Illumina, San Diego, CA) and amplified using ISOSDB412 IS-Seq Step1 and p7 primers for 13 cycles using Q5 Master Mix (New England Biolabs). The products of this reaction were amplified with IS-Seq Step2 and p7 primers for 9 cycles using Q5 Master Mix and sequenced on a Novaseq 6000 at Novogene (Sacramento, CA). R1 reads were binned and trimmed of the ISOSDB412 terminus and adapters using cutadapt v1.1.18 (110). To remove PCR duplicates, R2 pairs of R1 reads that contained the ISOSDB412 terminus were binned and deduplicated using dedupe.sh from the BBTools suite with the flag minidentity=100. Deduplication of R2 reads ensures that each amplified read pair is a unique molecule, while maintaining read depth information of the R1 (IS-amplified) read. This set of deduplicated R2 reads was re-paired to trimmed R1 reads with reformat.sh from the BBTools suite and these pairs were aligned to the host genome using Bowtie2. R1 read depth at each genomic position was calculated using bedtools v2.30.0 (111) IS seq reads can be found at the European Nucleotide Archive under study accession number PRJEB66483.

B. thetaiotaomicron passaging experiment

B. thetaiotaomicron was grown overnight in BHIS and then subcultured 1:100 into either Varel-Bryant minimal media or BHIS containing antibiotic and grown overnight. The next day, bacteria cultured in BHIS were subcultured 1:100 into BHIS and bacteria cultured in Varel-Bryant minimal media cultured were subcultured 1:100 into Varel-Bryant minimal media with antibiotic. After overnight growth, this passaging was repeated once more. Genomic DNA was isolated and the final passage was sequenced as described above.

B. thetaiotamicron phage infection

B. thetaiotamicron CPS3 (a mutant of *B. thetaiotamicron* VPI-5482 expressing only the CPS3 capsule) (62) (112) was transformed with pB6T-ermF-IS4351 and grown overnight in *Bacteroides* phage recovery medium (BPRM) with antibiotic (106). Overnight cultures were diluted to OD_{600} of 1.0 in phage buffer (62) and was plated on BPRM agar containing tetracycline and $1x10^8$ PFU/mL of either DAC15 or DAC17. Dilutions of resuspended overnight culture were also plated on BPRM containing tetracycline. Plated cultures on DAC15 agar media were incubated at 37° C for 48 hours and individual colonies were picked and patched onto fresh BPRM plates with DAC15. This process was repeated. Cultures plated on DAC17 agar and agar without phage were similarly passaged. Passaged colonies were cultured into BPRM broth without phage and incubated for 48 hours at 37° C. Genomic DNA was isolated and sequenced as described above.

Enumeration of infectious phage particles

90 µL of supernatant from liquid cultures of passaged colonies were mixed with 10 µL chloroform (Fisher Scientific), centrifuged for 1 minute at 21,000 RCF at 25° C, and the supernatant was collected. Dilutions of this supernatant were made in phage buffer and 5 µL

spots of the dilutions were plated on a BPRM plate containing *B. thetaiotamicron* CPS3 embedded in 0.35% BPRM top agar, and incubated at 37° C overnight.

Functional and taxonomic classification of contigs and ORFs

Metagenomic sequences with IS insertions were functionally annotated using eggNOG mapper v2 (113). Orthologous groups were cross-validated using the COG database (114). Antibiotic resistance genes were identified using CARD RGI 6.0.0 (115). ORFs and contigs containing IS insertions were classified using Kraken2 v2.0.7 (116).

Acknowledgements

We would like to thank the members of the Duerkop lab for helpful discussions related to this manuscript. We would like to thank A. Murat Eren and Alejandro Reyes Muñoz for critically reading this manuscript and providing useful insights.

Funding

This work was supported by the National Institutes of Health grants R01AI141479 (B.A.D.), T32AI052066 (J.M.K), F31AI169976 (J.M.K.), and R35GM150996 (A.J.H). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Figure Legends

Figure 1. The ISOSDB: a database of diverse and complete IS elements. (A) Number of unique IS element sequences in ISFinder and the ISOSDB. (B) Comparison of transposase (Tnp amino acid) and nucleotide sequences of the full IS elements from the ISOSDB to the IS elements from ISFinder. (C) Phylogenetic tree of the Tnp genes of IS elements in the ISOSDB. Leaves are colored by Tnp family.

Figure 2. IS element insertions are abundant in intestinal bacteria. (A) Abundance of IS elements from ITA, JPN, and MDG individuals. Each data point represents one new insertion not found in the reference assembly. The y-axis is the relative IS depth percentage (IS depth/ (WT allele depth + IS depth)). (B) Number of insertions per sample shown in A. (C) IS element insertions per bacterial class. The intensity of each bar is proportional to the number of IS element insertions. (D). IS element insertions organized by IS family that from insertions in *Bacteroidia* and *Clostridia* bacteria. The intensity of each bar is proportional to the number of IS element insertions. (E) Preferential IS element insertion in either intergenic (left side, red) or intragenic (right side, blue) loci for highly abundant IS elements (paired T test with FDR multiple testing correction, maximum adjusted p value = 0.05).

Figure 3. IS element insertions are commonly found in bacterial accessory genes. (A) Number of shared and unique iORFs within ITA, JPN, and MDG individuals. (B) Heatmap of IS element insertions in different gene functional categories. (C) Statistically significant enrichment of IS elements in the functional categories from B (Fisher's exact test with FDR multiple comparison correction, ***$p < 10^{-25}$, **$p < 10^{-4}$, *$p < 0.05$). (D) Location and relative IS element depth in shared iORFs. The location of the insertion (either above or below the gene representation) is representative of the study. (E-F) Number of IS element insertions in iORFs in the functional categories in B from different bacterial classes (E) and IS families (F). (G) Rate of IS insertions in transposases. Rectangles in the red dashed line are self-targets and rectangles outside of the red dashed line are non-synonymous-targets.

Figure 4. ISOSDB412 insertion in *Bacteroides* species replicates findings from intestinal metagenomic data. (A and D) IS-Seq deep sequencing of ISOSDB412 insertions in *Bt* (A) and *Bf* (D). (B and E) Functional enrichment of ISOSDB412 insertions in *Bt* (B) and *Bf* (E) (Fisher's

exact test, *$p < 0.06$, ***$p < 0.001$). (C and F) ISOSDB412 insertions were counted in intergenic and intragenic loci in *Bt* (C) and *Bf* (F) (paired T-test, *$p < 0.05$). (G) IS-Seq of ISOSDB412 insertions of *Bt* passaged in either BHIS or MM. (H) Number of ISOSDB412 insertions in each condition in G (paired T test). (I) Number of ISOSDB412 insertions in EPS biogenesis genes in G (paired T test). (J) IS-Seq of ISOSDB412 insertions of *Bt* chronically infected with DAC15 or DAC17. (K) Number of ISOSDB412 insertions in each condition in J (unpaired T test, *$p < 0.05$). (L) Number of ISOSDB412 insertions in EPS biogenesis genes in J (unpaired T test, *$p < 0.05$).

Figure 5. IS element insertions are maintained in the intestinal microbiota for extended time periods. (A) IS element insertions are depicted as rectangles, with the color intensity proportional to the relative IS depth percentage. Each insertion is unique based on insertion position and the same insertions are aligned vertically. Orange-colored IS element insertions are present at the first timepoint (week 0), green IS element insertions are new insertions not present at the initial timepoint, and blue IS element insertions are new insertions at timepoints after their first appearance. (B) Number of IS element insertions per timepoint for IS families (red) or taxonomic classes (blue). (C) Rate of new Is element insertions per week. (D) Dynamics of IS element insertions over time for gene functional category iORFs. Data from every individual was used for this figure.

Figure 6. FMT transfer of IS element insertions between donors and recipients. (A) Blue rectangles are IS element insertions in donor samples and red rectangles are IS element insertions in recipient samples. The same insertions in donor and recipients are aligned vertically. Week 0 is the transplanted fecal material. (B) Heatmap showing new IS element insertions that were detected 2 weeks after transplantation and are detected in both donor and recipient samples. (C) Loss of IS element insertions detected in the transplanted fecal material

(week 0) from recipients (paired T test, ***p* < 0.01). (D) Similar rates of new IS element insertions are found in both donor and recipients (unpaired T test).

Figure 7. Antibiotic treatment sanitizes IS element insertions from intestinal bacteria. (A) Heatmap showing IS element insertion dynamics during antibiotic treatment. Only insertions in alleles with less than or equal to 200% of the week 0 read depth are shown in this panel. Horizontal black bars at the 0.6 Week timepoint are present in individuals who did not have an available sample for this timepoint. (B) Rebounding and non-rebounding wild type alleles of IS element insertions lost from week 0 during antibiotic treatment. (C) Clostridia bacteria rapidly lose IS element insertions. Data from every individual was used for this figure panel. (D) Antibiotic treatment significantly decreases the number of IS element insertions present following week 0 (unpaired T test, *** *p* < 0.005).

Figure 8. Diet intervention alters IS element mediated *susCD/tonB* gene inactivation. (A) Heatmap showing IS element insertion dynamics before, during, and after diet intervention. Only insertions in alleles with less than or equal to 200% of the week 0 read depth are shown in this panel. (B) Rebounding and non-rebounding wild type alleles of IS element insertions lost from week 0 during diet intervention. (C) Dynamics of IS element insertions in *susCD/tonB* genes during the timecourse. Data from every individual was used for this figure panel. (D) New *BT1119* IS element insertions arise after the diet intervention has ended. Wild type allele depth is shown as a black line and IS element insertion allele depth is shown as a red bar. (E) Diet intervention significantly decreases the number of IS element insertions present following week 0 (unpaired T test, ****p* < 0.005).

Figure S1. Schematic of the pseudoR computational pipeline. Bacterial genomes are sequenced using Illumina short-read DNA sequencing. These reads are assembled into contigs

and the reads (black rectangles) are mapped against the assembly contigs. Unmapped reads (black and blue lines) are compared against the ISOSDB containing IS element termini. IS element termini are trimmed from the reads and the reads are then are re-mapped to the assembly contigs.

Figure S2. IS element insertions in diverse genera and functional classes. (A) Abundance of IS element insertions for intestinal bacteria at the genera level. (B) Replotting of the data in Fig. 3C to include the “other” category which encompasses all other genes containing IS element insertions that were not the focus of this study.

Figure S3. IS element insertions in genes that affect *Bt* fitness. (A) Table of IS element hits in genes involved in fitness in *Bt*. Genes whose fitness values < 0 are beneficial for fitness in the specified condition and genes whose fitness > 0 are detrimental to fitness in the specified condition (+++ fitness > 4, ++ fitness > 2, + fitness > 1, - - - fitness < -4, - - fitness < -2, - fitness < -1). (B) Schematic of IS element insertions in *BT1119*.

Figure S4. ISOSDB412 drives IS element insertions during selective pressure. (A) IS element insertions in *Bt* and *Bf* strains following passage in nutrient limited conditions (unpaired T test, $^{**}p < 0.01$) (B) Schematic of IS element insertions in BT3642. (C) Quantification of infectious phage particles in the supernatant of *Bt* CPS3 strains chronically infected with DAC15 or DAC17. (D) Schematic of IS element insertions in the CPS3 locus.

Figure S5-6. Remainder of healthy individuals followed longitudinally (Related to Fig. 5A-B). The same scale in 5A-B is used for these figures.

Figure S7. New IS element insertions are not associated with increased wild type allele depth. (A) The percentage of new IS element insertions is shown on the Y axis and the X axis is the percentage of the wild type allele depth from the first detection of each new insertion divided by the maximum detected wild type allele depth of the timepoints prior to the first detection timepoint. The red dashed line intersects the X axis at 200%. (B) Representation of abundance-driven or transposition-driven allele measurements. (C) Significantly more IS element insertions are predicted to arise from transposition compared to changes in host abundance (***$p < 0.01$, paired T-test). (D) Wild type allele depth of IS element insertions that were present at week 0 and were not detected at another timepoint. Each IS element insertion is shown as a line. Red lines ("rebounders") represent IS element insertions where the wild type allele depth at later timepoints reached at least 50% of the week 0 wild type allele depth. Black lines ("non-rebounders") represent IS element insertions where the wild type allele depth at later timepoints did not reached at least 50% of the week 0 wild type allele depth.

Figure S8. Remainder of individuals followed longitudinally after FMT (Related to Fig. 6A). Same scale in Fig. 6A is used in these figures.

Figure S9. Remainder of shared insertion plots (Related to Fig. 6B).

Figure S10. Remainder of antibiotic treated individual plots (Related to Fig. 7A and 7B).

Figure S11. Remainder of antibiotic treated individual plots (Related to Fig. 8A and 8B).

References

1. Shkoporov AN, Clooney AG, Sutton TDS, Ryan FJ, Daly KM, Nolan JA, McDonnell SA, Khokhlova EV, Draper LA, Forde A, Guerin E, Velayudhan V, Ross RP, Hill C. 2019.

The human gut virome is highly diverse, stable, and individual specific. Cell Host Microbe 26:527-541.e5. 10.1016/j.chom.2019.09.009

2. Forster SC, Liu J, Kumar N, Gulliver EL, Gould JA, Escobar-Zepeda A, Mkandawire T, Pike LJ, Shao Y, Stares MD, Browne HP, Neville BA, Lawley TD. 2022. Strain-level characterization of broad host range mobile genetic elements transferring antibiotic resistance from the human microbiome. Nat Commun 13:1445. 10.1038/s41467-022-29096-9

3. Jiang X, Hall AB, Xavier RJ, Alm EJ. 2019. Comprehensive analysis of chromosomal mobile genetic elements in the gut microbiome reveals phylum-level niche-adaptive gene pools. PLoS One 14:e0223680. 10.1371/journal.pone.0223680

4. Jiang X, Hall AB, Arthur TD, Plichta DR, Covington CT, Poyet M, Crothers J, Moses PL, Tolonen AC, Vlamakis H, Alm EJ, Xavier RJ. 2019. Invertible promoters mediate bacterial phase variation, antibiotic resistance, and host adaptation in the gut. Science 363:181-187. 10.1126/science.aau5238

5. Desvaux M, Dalmasso G, Beyrouthy R, Barnich N, Delmas J, Bonnet R. 2020. Pathogenicity factors of genomic islands in intestinal and extraintestinal *Escherichia coli*. Front Microbiol 11:2065. 10.3389/fmicb.2020.02065

6. Groussin M, Poyet M, Sistiaga A, Kearney SM, Moniz K, Noel M, Hooker J, Gibbons SM, Segurel L, Froment A, Mohamed RS, Fezeu A, Juimo VA, Lafosse S, Tabe FE, Girard C, Iqaluk D, Nguyen LTT, Shapiro BJ, Lehtimäki J, Ruokolainen L, Kettunen PP, Vatanen T, Sigwazi S, Mabulla A, Domínguez-Rodrigo M, Nartey YA, Agyei-Nkansah A, Duah A, Awuku YA, Valles KA, Asibey SO, Afihene MY, Roberts LR, Plymoth A, Onyekwere CA, Summons RE, Xavier RJ, Alm EJ. 2021. Elevated rates of horizontal

gene transfer in the industrialized human microbiome. Cell 184:2053-2067.e18. 10.1016/j.cell.2021.02.052

7. Brito IL, Yilmaz S, Huang K, Xu L, Jupiter SD, Jenkins AP, Naisilisili W, Tamminen M, Smillie CS, Wortman JR, Birren BW, Xavier RJ, Blainey PC, Singh AK, Gevers D, Alm EJ. 2016. Mobile genes in the human microbiome are structured from global to individual scales. Nature 535:435-439. 10.1038/nature18927

8. Aziz RK, Breitbart M, Edwards RA. 2010. Transposases are the most abundant, most ubiquitous genes in nature. Nucleic Acids Res 38:4207-17. 10.1093/nar/gkq140

9. Siguier P, Gourbeyre E, Chandler M. 2014. Bacterial insertion sequences: their genomic impact and diversity. FEMS Microbiol Rev 38:865-91. 10.1111/1574-6976.12067

10. He S, Corneloup A, Guynet C, Lavatine L, Caumont-Sarcos A, Siguier P, Marty B, Dyda F, Chandler M, Ton Hoang B. 2015. The IS200/IS605 family and "peel and paste" single-strand transposition mechanism. Microbiol Spectr 3. 10.1128/microbiolspec.MDNA3-0039-2014

11. Piegu B, Bire S, Arensburger P, Bigot Y. 2015. A survey of transposable element classification systems--a call for a fundamental update to meet the challenge of their diversity and complexity. Mol Phylogenet Evol 86:90-109. 10.1016/j.ympev.2015.03.009

12. Kosek D, Hickman AB, Ghirlando R, He S, Dyda F. 2021. Structures of ISCth4 transpososomes reveal the role of asymmetry in copy-out/paste-in DNA transposition. EMBO J 40:e105666. 10.15252/embj.2020105666

13. Jellen-Ritter AS, Kern WV. 2001. Enhanced expression of the multidrug efflux pumps AcrAB and AcrEF associated with insertion element transposition in *Escherichia coli*

mutants selected with a fluoroquinolone. Antimicrob Agents Chemother 45:1467-72. 10.1128/aac.45.5.1467-1472.2001

14. Shankar N, Baghdayan AS, Gilmore MS. 2002. Modulation of virulence within a pathogenicity island in vancomycin-resistant *Enterococcus faecalis*. Nature 417:746-50. 10.1038/nature00802

15. Chain PS, Carniel E, Larimer FW, Lamerdin J, Stoutland PO, Regala WM, Georgescu AM, Vergez LM, Land ML, Motin VL, Brubaker RR, Fowler J, Hinnebusch J, Marceau M, Medigue C, Simonet M, Chenal-Francisque V, Souza B, Dacheux D, Elliott JM, Derbise A, Hauser LJ, Garcia E. 2004. Insights into the evolution of *Yersinia pestis* through whole-genome comparison with *Yersinia pseudotuberculosis*. Proc Natl Acad Sci U S A 101:13826-31. 10.1073/pnas.0404012101

16. Manson JM, Hancock LE, Gilmore MS. 2010. Mechanism of chromosomal transfer of *Enterococcus faecalis* pathogenicity island, capsule, antimicrobial resistance, and other traits. Proc Natl Acad Sci U S A 107:12269-74. 10.1073/pnas.1000139107

17. Hawkey J, Monk JM, Billman-Jacobe H, Palsson B, Holt KE. 2020. Impact of insertion sequences on convergent evolution of *Shigella* species. PLoS Genet 16:e1008931. 10.1371/journal.pgen.1008931

18. Leavis HL, Willems RJ, van Wamel WJ, Schuren FH, Caspers MP, Bonten MJ. 2007. Insertion sequence-driven diversification creates a globally dispersed emerging multiresistant subspecies of *E. faecium*. PLoS Pathog 3:e7. 10.1371/journal.ppat.0030007

19. Kirsch JM, Ely S, Stellfox ME, Hullahalli K, Luong P, Palmer KL, Van Tyne D, Duerkop BA. 2023. Targeted IS-element sequencing uncovers transposition dynamics during

selective pressure in enterococci. PLoS Pathog 19:e1011424. 10.1371/journal.ppat.1011424

20. Zlitni S, Bishara A, Moss EL, Tkachenko E, Kang JB, Culver RN, Andermann TM, Weng Z, Wood C, Handy C, Ji HP, Batzoglou S, Bhatt AS. 2020. Strain-resolved microbiome sequencing reveals mobile elements that drive bacterial competition on a clinical timescale. Genome Med 12:50. 10.1186/s13073-020-00747-0

21. Zhao S, Lieberman TD, Poyet M, Kauffman KM, Gibbons SM, Groussin M, Xavier RJ, Alm EJ. 2019. Adaptive evolution within gut microbiomes of healthy people. Cell Host Microbe 25:656-667.e8. 10.1016/j.chom.2019.03.007

22. Elhenawy W, Tsai CN, Coombes BK. 2019. Host-specific adaptive diversification of Crohn's disease-associated adherent-invasive *Escherichia coli*. Cell Host Microbe 25:301-312.e5. 10.1016/j.chom.2018.12.010

23. Barreto HC, Frazão N, Sousa A, Konrad A, Gordo I. 2020. Mutation accumulation and horizontal gene transfer in *Escherichia coli* colonizing the gut of old mice. Commun Integr Biol 13:89-96. 10.1080/19420889.2020.1783059

24. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res 34:D32-6. 10.1093/nar/gkj014

25. Kichenaradja P, Siguier P, Pérochon J, Chandler M. 2010. ISbrowser: an extension of ISfinder for visualizing insertion sequences in prokaryotic genomes. Nucleic Acids Res 38:D62-8. 10.1093/nar/gkp947

26. Gounot JS, Chia M, Bertrand D, Saw WY, Ravikrishnan A, Low A, Ding Y, Ng AHQ, Tan LWL, Teo YY, Seedorf H, Nagarajan N. 2022. Genome-centric analysis of short and long read metagenomes reveals uncharacterized microbiome diversity in Southeast Asians. Nat Commun 13:6044. 10.1038/s41467-022-33782-z

27. Robinson DG, Lee MC, Marx CJ. 2012. OASIS: an automated program for global investigation of bacterial and archaeal insertion sequences. Nucleic Acids Res 40:e174. 10.1093/nar/gks778

28. Tempel S, Bedo J, Talla E. 2022. From a large-scale genomic analysis of insertion sequences to insights into their regulatory roles in prokaryotes. BMC Genomics 23:451. 10.1186/s12864-022-08678-3

29. Nzabarushimana E, Tang H. 2018. Insertion sequence elements-mediated structural variations in bacterial genomes. Mob DNA 9:29. 10.1186/s13100-018-0134-3

30. Quadrana L, Bortolini Silveira A, Mayhew GF, LeBlanc C, Martienssen RA, Jeddeloh JA, Colot V. 2016. The *Arabidopsis thaliana* mobilome and its impact at the species level. Elife 5:e15716. 10.7554/eLife.15716

31. Hawkey J, Hamidian M, Wick RR, Edwards DJ, Billman-Jacobe H, Hall RM, Holt KE. 2015. ISMapper: identifying transposase insertion sites in bacterial genomes from short read sequence data. BMC Genomics 16:667. 10.1186/s12864-015-1860-2

32. Durrant MG, Li MM, Siranosian BA, Montgomery SB, Bhatt AS. 2020. A bioinformatic analysis of integrative mobile genetic elements highlights their role in bacterial adaptation. Cell Host Microbe 27:140-153.e9. 10.1016/j.chom.2019.10.022

33. Sheng Y, Wang H, Ou Y, Wu Y, Ding W, Tao M, Lin S, Deng Z, Bai L, Kang Q. 2023. Insertion sequence transposition inactivates CRISPR-Cas immunity. Nat Commun 14:4366. 10.1038/s41467-023-39964-7

34. Rampelli S, Soverini M, D'Amico F, Barone M, Tavella T, Monti D, Capri M, Astolfi A, Brigidi P, Biagi E, Franceschi C, Turroni S, Candela M. 2020. Shotgun metagenomics of gut microbiota in humans with up to extreme longevity and the increasing role of xenobiotic degradation. mSystems 5. 10.1128/mSystems.00124-20

35. Yachida S, Mizutani S, Shiroma H, Shiba S, Nakajima T, Sakamoto T, Watanabe H, Masuda K, Nishimoto Y, Kubo M, Hosoda F, Rokutan H, Matsumoto M, Takamaru H, Yamada M, Matsuda T, Iwasaki M, Yamaji T, Yachida T, Soga T, Kurokawa K, Toyoda A, Ogura Y, Hayashi T, Hatakeyama M, Nakagama H, Saito Y, Fukuda S, Shibata T, Yamada T. 2019. Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. Nat Med 25:968-976. 10.1038/s41591-019-0458-7

36. Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, Armanini F, Beghini F, Manghi P, Tett A, Ghensi P, Collado MC, Rice BL, DuLong C, Morgan XC, Golden CD, Quince C, Huttenhower C, Segata N. 2019. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. Cell 176:649-662.e20. 10.1016/j.cell.2019.01.001

37. Rinninella E, Raoul P, Cintoni M, Franceschi F, Miggiano GAD, Gasbarrini A, Mele MC. 2019. What is the healthy gut microbiota composition? A changing ecosystem across age, environment, diet, and diseases. Microorganisms 7. 10.3390/microorganisms7010014

38. Alpizar-Rodriguez D, Lesker TR, Gronow A, Gilbert B, Raemy E, Lamacchia C, Gabay C, Finckh A, Strowig T. 2019. *Prevotella copri* in individuals at risk for rheumatoid arthritis. Ann Rheum Dis 78:590-593. 10.1136/annrheumdis-2018-214514

39. Scher JU, Sczesnak A, Longman RS, Segata N, Ubeda C, Bielski C, Rostron T, Cerundolo V, Pamer EG, Abramson SB, Huttenhower C, Littman DR. 2013. Expansion of intestinal *Prevotella copri* correlates with enhanced susceptibility to arthritis. Elife 2:e01202. 10.7554/eLife.01202

40. Hughes ER, Winter MG, Duerkop BA, Spiga L, Furtado de Carvalho T, Zhu W, Gillis CC, Buttner L, Smoot MP, Behrendt CL, Cherry S, Santos RL, Hooper LV, Winter SE. 2017. Microbial respiration and formate oxidation as metabolic signatures of inflammation-associated dysbiosis. Cell Host Microbe 21:208-219. 10.1016/j.chom.2017.01.005

41. Conway T, Cohen PS. 2015. Commensal and pathogenic *Escherichia coli* metabolism in the gut. Microbiol Spectr 3. 10.1128/microbiolspec.MBP-0006-2014

42. Lebreton F, van Schaik W, McGuire AM, Godfrey P, Griggs A, Mazumdar V, Corander J, Cheng L, Saif S, Young S, Zeng Q, Wortman J, Birren B, Willems RJ, Earl AM, Gilmore MS. 2013. Emergence of epidemic multidrug-resistant *Enterococcus faecium* from animal and commensal strains. mBio 4. 10.1128/mBio.00534-13

43. McEvoy CR, Tsuji B, Gao W, Seemann T, Porter JL, Doig K, Ngo D, Howden BP, Stinear TP. 2013. Decreased vancomycin susceptibility in *Staphylococcus aureus* caused by IS256 tempering of WalKR expression. Antimicrob Agents Chemother 57:3240-9. 10.1128/AAC.00279-13

44. Kozitskaya S, Cho SH, Dietrich K, Marre R, Naber K, Ziebuhr W. 2004. The bacterial insertion sequence element IS256 occurs preferentially in nosocomial *Staphylococcus*

epidermidis isolates: association with biofilm formation and resistance to aminoglycosides. Infect Immun 72:1210-5. 10.1128/IAI.72.2.1210-1215.2004

45. Martens EC, Koropatkin NM, Smith TJ, Gordon JI. 2009. Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. J Biol Chem 284:24673-7. 10.1074/jbc.R109.022848

46. Noinaj N, Guillier M, Barnard TJ, Buchanan SK. 2010. TonB-dependent transporters: regulation, structure, and function. Annu Rev Microbiol 64:43-60. 10.1146/annurev.micro.112408.134247

47. Yadav AK, Espaillat A, Cava F. 2018. Bacterial Strategies to Preserve Cell Wall Integrity Against Environmental Threats. Front Microbiol 9:2064. 10.3389/fmicb.2018.02064

48. Vassallo CN, Doering CR, Littlehale ML, Teodoro GIC, Laub MT. 2022. A functional selection reveals previously undetected anti-phage defence systems in the *E. coli* pangenome. Nat Microbiol 7:1568-1579. 10.1038/s41564-022-01219-4

49. Aravind L, Iyer LM, Leipe DD, Koonin EV. 2004. A novel family of P-loop NTPases with an unusual phyletic distribution and transmembrane segments inserted within the NTPase domain. Genome Biol 5:R30. 10.1186/gb-2004-5-5-r30

50. Komano T. 1999. Shufflons: multiple inversion systems and integrons. Annu Rev Genet 33:171-91. 10.1146/annurev.genet.33.1.171

51. Blattner FR, Plunkett G, 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y. 1997. The complete genome sequence of *Escherichia coli* K-12. Science 277:1453-62. 10.1126/science.277.5331.1453

52. Lyon BR, Gillespie MT, Skurray RA. 1987. Detection and characterization of IS256, an insertion sequence in *Staphylococcus aureus*. J Gen Microbiol 133:3031-8. 10.1099/00221287-133-11-3031

53. Quintiliani R, Jr., Courvalin P. 1996. Characterization of Tn1547, a composite transposon flanked by the IS16 and IS256-like elements, that confers vancomycin resistance in *Enterococcus faecalis* BM4281. Gene 172:1-8. 10.1016/0378-1119(96)00110-2

54. Dufour A, Rince A, Uguen P, Le Pennec JP. 2000. IS1675, a novel lactococcal insertion element, forms a transposon-like structure including the lacticin 481 lantibiotic operon. J Bacteriol 182:5600-5. 10.1128/JB.182.19.5600-5605.2000

55. Udaondo Z, Abram KZ, Kothari A, Jun S-R. 2021. Insertion sequences associated with antibiotic resistance genes in *Enterococcus* isolates from an inpatient with prolonged bacteremia. bioRxiv doi:10.1101/2021.07.22.453447. 10.1101/2021.07.22.453447

56. Liu H, Shiver AL, Price MN, Carlson HK, Trotter VV, Chen Y, Escalante V, Ray J, Hern KE, Petzold CJ, Turnbaugh PJ, Huang KC, Arkin AP, Deutschbauer AM. 2021. Functional genetics of human gut commensal *Bacteroides thetaiotaomicron* reveals metabolic requirements for growth across environments. Cell Rep 34:108789. 10.1016/j.celrep.2021.108789

57. Flowers SA, Evans SJ, Ward KM, McInnis MG, Ellingrod VL. 2017. Interaction between atypical antipsychotics and the gut microbiome in a bipolar disease cohort. Pharmacotherapy 37:261-267. 10.1002/phar.1890

58. Maier L, Pruteanu M, Kuhn M, Zeller G, Telzerow A, Anderson EE, Brochado AR, Fernandez KC, Dose H, Mori H, Patil KR, Bork P, Typas A. 2018. Extensive impact of non-antibiotic drugs on human gut bacteria. Nature 555:623-628. 10.1038/nature25979

59. Ait Chait Y, Mottawea W, Tompkins TA, Hammami R. 2020. Unravelling the antimicrobial action of antidepressants on gut commensal microbes. Sci Rep 10:17878. 10.1038/s41598-020-74934-9

60. Kyono Y, Ellezian L, Hu Y, Eliadis K, Moy J, Hirsch EB, Federle MJ, Flowers SA. 2022. The atypical antipsychotic quetiapine promotes multiple antibiotic resistance in *Escherichia coli*. J Bacteriol 204:e0010222. 10.1128/jb.00102-22

61. Dutilh BE, Cassman N, McNair K, Sanchez SE, Silva GG, Boling L, Barr JJ, Speth DR, Seguritan V, Aziz RK, Felts B, Dinsdale EA, Mokili JL, Edwards RA. 2014. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. Nat Commun 5:4498. 10.1038/ncomms5498

62. Hryckowian AJ, Merrill BD, Porter NT, Van Treuren W, Nelson EJ, Garlena RA, Russell DA, Martens EC, Sonnenburg JL. 2020. Bacteroides thetaiotaomicron-Infecting Bacteriophage Isolates Inform Sequence-Based Host Range Predictions. Cell Host Microbe 28:371-379.e5. 10.1016/j.chom.2020.06.011

63. Watson AR, Füssel J, Veseli I, DeLongchamp JZ, Silva M, Trigodet F, Lolans K, Shaiber A, Fogarty E, Runde JM, Quince C, Yu MK, Söylev A, Morrison HG, Lee STM, Kao D, Rubin DT, Jabri B, Louie T, Eren AM. 2023. Metabolic independence drives gut microbial colonization and resilience in health and disease. Genome Biol 24:78. 10.1186/s13059-023-02924-x

64. Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, Andrews E, Ajami NJ, Bonham KS, Brislawn CJ, Casero D, Courtney H, Gonzalez A, Graeber TG, Hall AB, Lake K, Landers CJ, Mallick H, Plichta DR, Prasad M, Rahnavard G, Sauk J, Shungin D, Vázquez-Baeza Y, White RA, 3rd, Braun J, Denson LA, Jansson JK, Knight R, Kugathasan S, McGovern DPB, Petrosino JF, Stappenbeck TS, Winter HS, Clish CB, Franzosa EA, Vlamakis H, Xavier RJ, Huttenhower C. 2019. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature 569:655-662. 10.1038/s41586-019-1237-9

65. Palleja A, Mikkelsen KH, Forslund SK, Kashani A, Allin KH, Nielsen T, Hansen TH, Liang S, Feng Q, Zhang C, Pyl PT, Coelho LP, Yang H, Wang J, Typas A, Nielsen MF, Nielsen HB, Bork P, Wang J, Vilsbøll T, Hansen T, Knop FK, Arumugam M, Pedersen O. 2018. Recovery of gut microbiota of healthy adults following antibiotic exposure. Nat Microbiol 3:1255-1265. 10.1038/s41564-018-0257-9

66. Louis S, Tappu RM, Damms-Machado A, Huson DH, Bischoff SC. 2016. Characterization of the gut microbial community of obese patients following a weight-loss intervention using whole metagenome shotgun sequencing. PLoS One 11:e0149564. 10.1371/journal.pone.0149564

67. Bischoff SC, Damms-Machado A, Betz C, Herpertz S, Legenbauer T, Löw T, Wechsler JG, Bischoff G, Austel A, Ellrott T. 2012. Multicenter evaluation of an interdisciplinary 52-week weight loss program for obesity with regard to body weight, comorbidities and quality of life--a prospective study. Int J Obes (Lond) 36:614-24. 10.1038/ijo.2011.107

68. Schloissnig S, Arumugam M, Sunagawa S, Mitreva M, Tap J, Zhu A, Waller A, Mende DR, Kultima JR, Martin J, Kota K, Sunyaev SR, Weinstock GM, Bork P. 2013. Genomic

variation landscape of the human gut microbiome. Nature 493:45-50. 10.1038/nature11711

69. Chanin RB, West PT, Park RM, Wirbel J, Green GZM, Miklos AM, Gill MO, Hickey AS, Brooks EF, Bhatt AS. 2023. Intragenic DNA inversions expand bacterial coding capacity. bioRxiv doi:10.1101/2023.03.11.532203. 10.1101/2023.03.11.532203

70. Milman O, Yelin I, Kishony R. 2023. Systematic identification of gene-altering programmed inversions across the bacterial domain. Nucleic Acids Res 51:553-573. 10.1093/nar/gkac1166

71. Pedersen TK, Brown EM, Plichta DR, Johansen J, Twardus SW, Delorey TM, Lau H, Vlamakis H, Moon JJ, Xavier RJ, Graham DB. 2022. The CD4(+) T cell response to a commensal-derived epitope transitions from a tolerant to an inflammatory state in Crohn's disease. Immunity 55:1909-1923.e6. 10.1016/j.immuni.2022.08.016

72. Curtis MM, Hu Z, Klimko C, Narayanan S, Deberardinis R, Sperandio V. 2014. The gut commensal *Bacteroides thetaiotaomicron* exacerbates enteric infection through modification of the metabolic landscape. Cell Host Microbe 16:759-69. 10.1016/j.chom.2014.11.005

73. Spiga L, Winter MG, Furtado de Carvalho T, Zhu W, Hughes ER, Gillis CC, Behrendt CL, Kim J, Chessa D, Andrews-Polymenis HL, Beiting DP, Santos RL, Hooper LV, Winter SE. 2017. An oxidative central metabolism enables *Salmonella* to utilize microbiota-derived succinate. Cell Host Microbe 22:291-301 e6. 10.1016/j.chom.2017.07.018

74. Jimenez AG, Ellermann M, Abbott W, Sperandio V. 2020. Diet-derived galacturonic acid regulates virulence and intestinal colonization in enterohaemorrhagic *Escherichia coli* and *Citrobacter rodentium*. Nat Microbiol 5:368-378. 10.1038/s41564-019-0641-0

75. De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poullet JB, Massart S, Collini S, Pieraccini G, Lionetti P. 2010. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. Proc Natl Acad Sci U S A 107:14691-6. 10.1073/pnas.1005963107

76. Golden CD, Anjaranirina EJG, Fernald LCH, Hartl DL, Kremen C, Milner DA, Jr., Ralalason DH, Ramihantaniarivo H, Randriamady H, Rice BL, Vaitla B, Volkman SK, Vonona MA, Myers SS. 2017. Cohort Profile: The Madagascar Health and Environmental Research (MAHERY) study in north-eastern Madagascar. Int J Epidemiol 46:1747-1748d. 10.1093/ije/dyx071

77. Carter MM, Olm MR, Merrill BD, Dahan D, Tripathi S, Spencer SP, Yu FB, Jain S, Neff N, Jha AR, Sonnenburg ED, Sonnenburg JL. 2023. Ultra-deep sequencing of Hadza hunter-gatherers recovers vanishing gut microbes. Cell 186:3111-3124.e13. 10.1016/j.cell.2023.05.046

78. Rasmussen JL, Odelson DA, Macrina FL. 1986. Complete nucleotide sequence and transcription of *ermF*, a macrolide-lincosamide-streptogramin B resistance determinant from *Bacteroides fragilis*. J Bacteriol 168:523-33. 10.1128/jb.168.2.523-533.1986

79. Garud NR, Good BH, Hallatschek O, Pollard KS. 2019. Evolutionary dynamics of bacteria in the gut microbiome within and across hosts. PLoS Biol 17:e3000102. 10.1371/journal.pbio.3000102

80. Wilson BC, Vatanen T, Jayasinghe TN, Leong KSW, Derraik JGB, Albert BB, Chiavaroli V, Svirskis DM, Beck KL, Conlon CA, Jiang Y, Schierding W, Holland DJ, Cutfield WS, O'Sullivan JM. 2021. Strain engraftment competition and functional augmentation in a multi-donor fecal microbiota transplantation trial for obesity. Microbiome 9:107. 10.1186/s40168-021-01060-7

81. Li SS, Zhu A, Benes V, Costea PI, Hercog R, Hildebrand F, Huerta-Cepas J, Nieuwdorp M, Salojärvi J, Voigt AY, Zeller G, Sunagawa S, de Vos WM, Bork P. 2016. Durable coexistence of donor and recipient strains after fecal microbiota transplantation. Science 352:586-9. 10.1126/science.aad8852

82. Smillie CS, Sauk J, Gevers D, Friedman J, Sung J, Youngster I, Hohmann EL, Staley C, Khoruts A, Sadowsky MJ, Allegretti JR, Smith MB, Xavier RJ, Alm EJ. 2018. Strain tracking reveals the determinants of bacterial engraftment in the human gut following fecal microbiota transplantation. Cell Host Microbe 23:229-240.e5. 10.1016/j.chom.2018.01.003

83. Ng KM, Aranda-Díaz A, Tropini C, Frankel MR, Van Treuren W, O'Loughlin CT, Merrill BD, Yu FB, Pruss KM, Oliveira RA, Higginbottom SK, Neff NF, Fischbach MA, Xavier KB, Sonnenburg JL, Huang KC. 2019. Recovery of the gut microbiota after antibiotics depends on host diet, community context, and environmental reservoirs. Cell Host Microbe 26:650-665.e4. 10.1016/j.chom.2019.10.011

84. Roodgar M, Good BH, Garud NR, Martis S, Avula M, Zhou W, Lancaster SM, Lee H, Babveyh A, Nesamoney S, Pollard KS, Snyder MP. 2021. Longitudinal linked-read sequencing reveals ecological and evolutionary responses of a human gut microbiome during antibiotic treatment. Genome Res 31:1433-1446. 10.1101/gr.265058.120

85. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling AV, Devlin AS, Varma Y, Fischbach MA, Biddinger SB, Dutton RJ, Turnbaugh PJ. 2014. Diet rapidly and reproducibly alters the human gut microbiome. Nature 505:559-63. 10.1038/nature12820

86. Leeming ER, Johnson AJ, Spector TD, Le Roy CI. 2019. Effect of diet on the gut microbiota: Rethinking intervention duration. Nutrients 11. 10.3390/nu11122862

87. Nagel M, Reuter T, Jansen A, Szekat C, Bierbaum G. 2011. Influence of ciprofloxacin and vancomycin on mutation rate and transposition of IS256 in *Staphylococcus aureus*. Int J Med Microbiol 301:229-36. 10.1016/j.ijmm.2010.08.021

88. Truong DT, Tett A, Pasolli E, Huttenhower C, Segata N. 2017. Microbial strain-level population structure and genetic diversity from metagenomes. Genome Res 27:626-638. 10.1101/gr.216242.116

89. Lee M. 2022. bit: a multipurpose collection of bioinformatics tools. F1000Research 11:122. 10.5281/zenodo.3383647

90. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068-9. 10.1093/bioinformatics/btu153

91. Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658-9. 10.1093/bioinformatics/btl158

92. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403-10. 10.1016/s0022-2836(05)80360-2

93. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772-80. 10.1093/molbev/mst010

94. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490. 10.1371/journal.pone.0009490

95. Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49:W293-w296. 10.1093/nar/gkab301

96. Leinonen R, Sugawara H, Shumway M. 2011. The sequence read archive. Nucleic Acids Res 39:D19-21. 10.1093/nar/gkq1019

97. Bushnell B. BBTools, SourceForge. https://sourceforge.net/projects/bbmap/.

98. Li D, Liu CM, Luo R, Sadakane K, Lam TW. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31:1674-6. 10.1093/bioinformatics/btv033

99. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. 10.1186/1471-2105-11-119

100. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357-9. 10.1038/nmeth.1923

101. Pedersen BS, Quinlan AR. 2018. Mosdepth: quick coverage calculation for genomes and exomes. Bioinformatics 34:867-868. 10.1093/bioinformatics/btx699

102. Shen W, Le S, Li Y, Hu F. 2016. SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. PLoS One 11:e0163962. 10.1371/journal.pone.0163962

103. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078-9. 10.1093/bioinformatics/btp352

104. Bacic MK, Smith CJ. 2008. Laboratory maintenance and cultivation of *Bacteroides* species. Curr Protoc Microbiol Chapter 13:Unit 13C.1. 10.1002/9780471729259.mc13c01s9

105. Varel VH, Bryant MP. 1974. Nutritional features of *Bacteroides fragilis* subsp. *fragilis*. Appl Microbiol 28:251-7. 10.1128/am.28.2.251-257.1974

106. Tartera C, Araujo R, Michel T, Jofre J. 1992. Culture and decontamination methods affecting enumeration of phages infecting Bacteroides fragilis in sewage. Appl Environ Microbiol 58:2670-3. 10.1128/aem.58.8.2670-2673.1992

107. Zheng L, Tan Y, Hu Y, Shen J, Qu Z, Chen X, Ho CL, Leung EL, Zhao W, Dai L. 2022. CRISPR/Cas-Based Genome Editing for Human Gut Commensal Bacteroides Species. ACS Synth Biol 11:464-472. 10.1021/acssynbio.1c00543

108. Willett JLE, Ji MM, Dunny GM. 2019. Exploiting biofilm phenotypes for functional characterization of hypothetical genes in Enterococcus faecalis. NPJ Biofilms Microbiomes 5:23. 10.1038/s41522-019-0099-0

109. Jones KR, Belvin BR, Macrina FL, Lewis JP. 2020. Sequence and characterization of shuttle vectors for molecular cloning in Porphyromonas, Bacteroides and related bacteria. Mol Oral Microbiol 35:181-191. 10.1111/omi.12304

110. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal 17:3. 10.14806/ej.17.1.200

111. Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841-2. 10.1093/bioinformatics/btq033

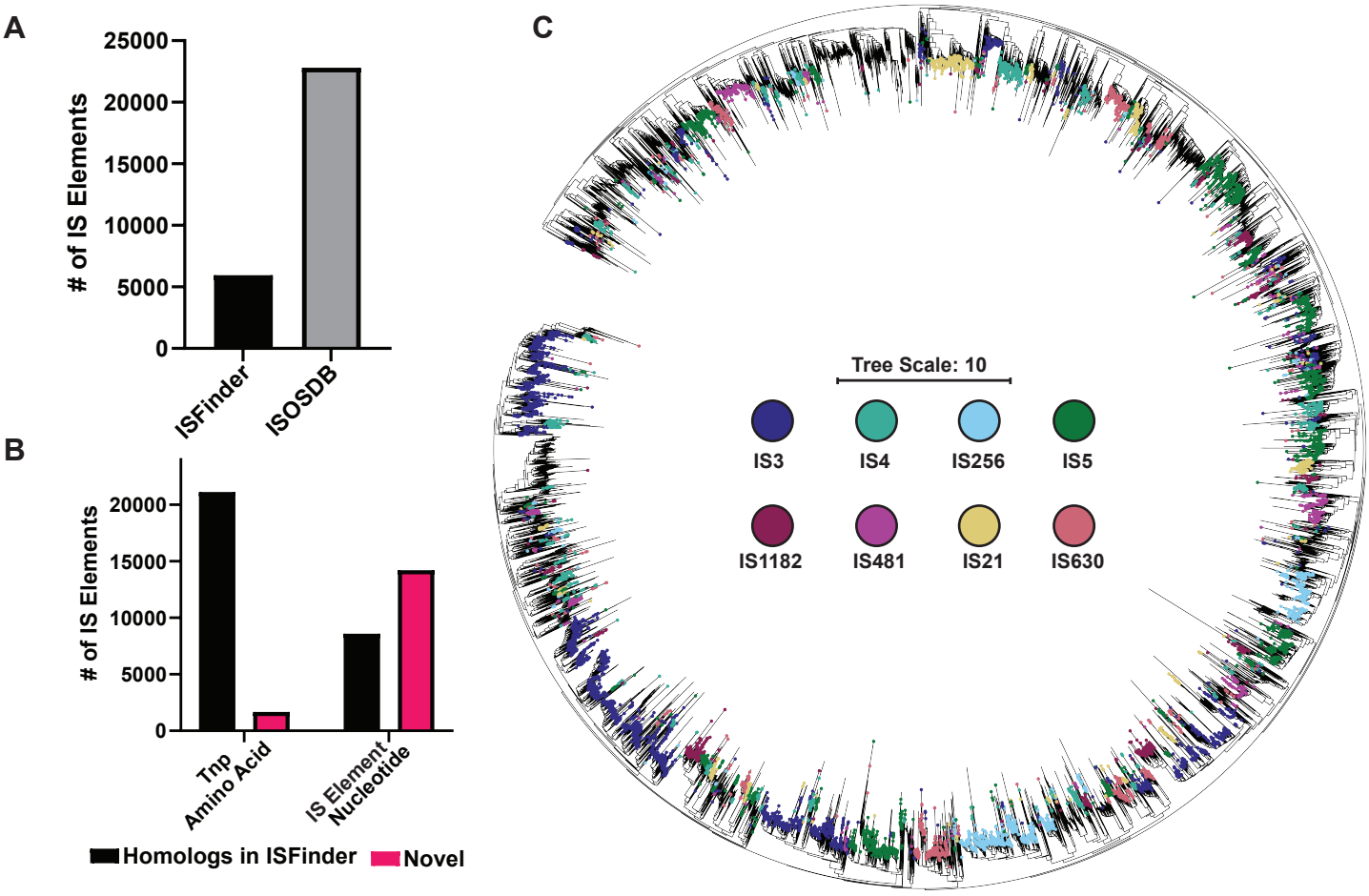
112. Hickey CA, Kuhn KA, Donermeyer DL, Porter NT, Jin C, Cameron EA, Jung H, Kaiko GE, Wegorzewska M, Malvin NP, Glowacki RW, Hansson GC, Allen PM, Martens EC, Stappenbeck TS. 2015. Colitogenic Bacteroides thetaiotaomicron Antigens Access Host Immune Cells in a Sulfatase-Dependent Manner via Outer Membrane Vesicles. Cell Host Microbe 17:672-80. 10.1016/j.chom.2015.04.002

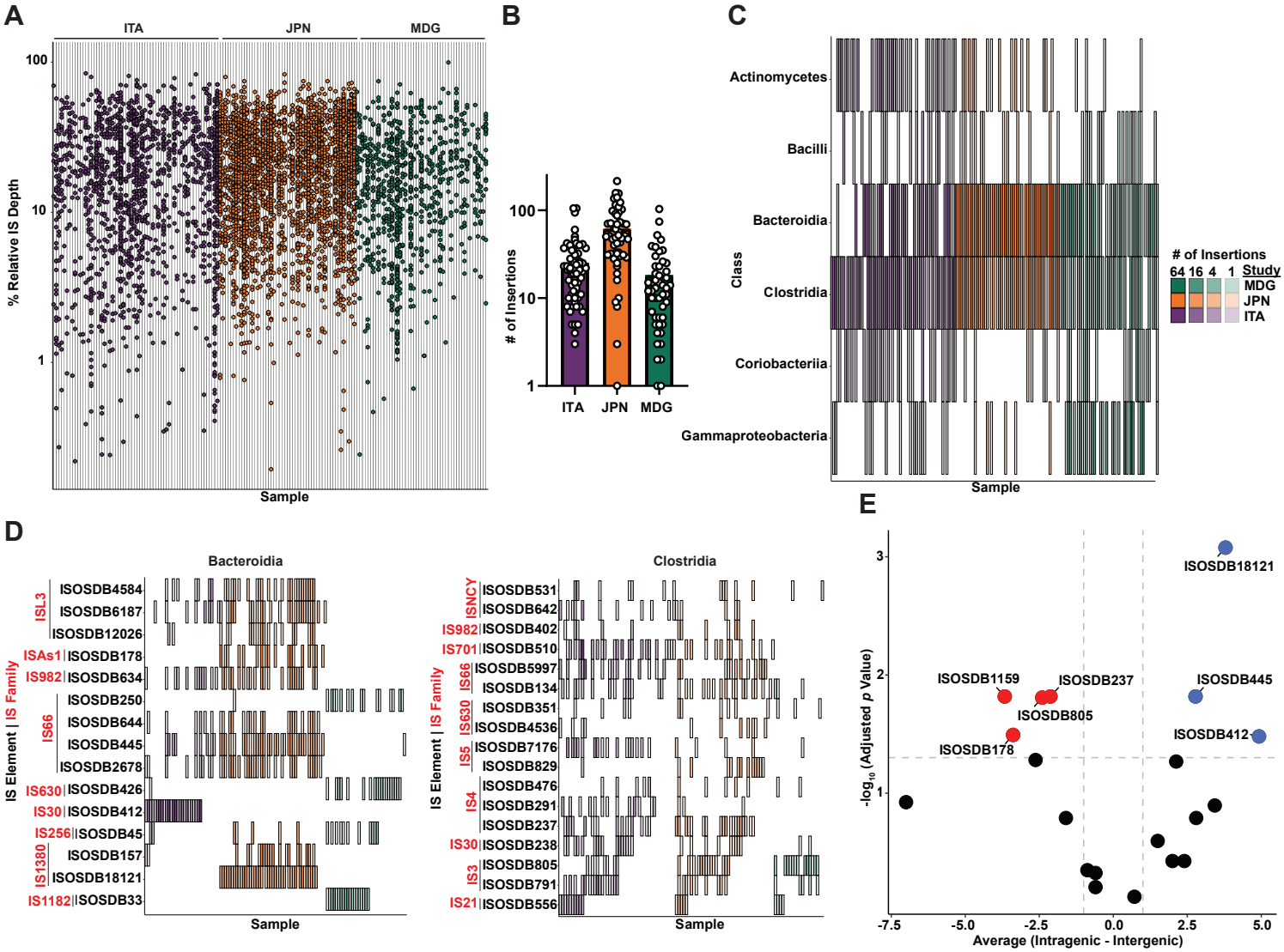
113. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol 38:5825-5829. 10.1093/molbev/msab293

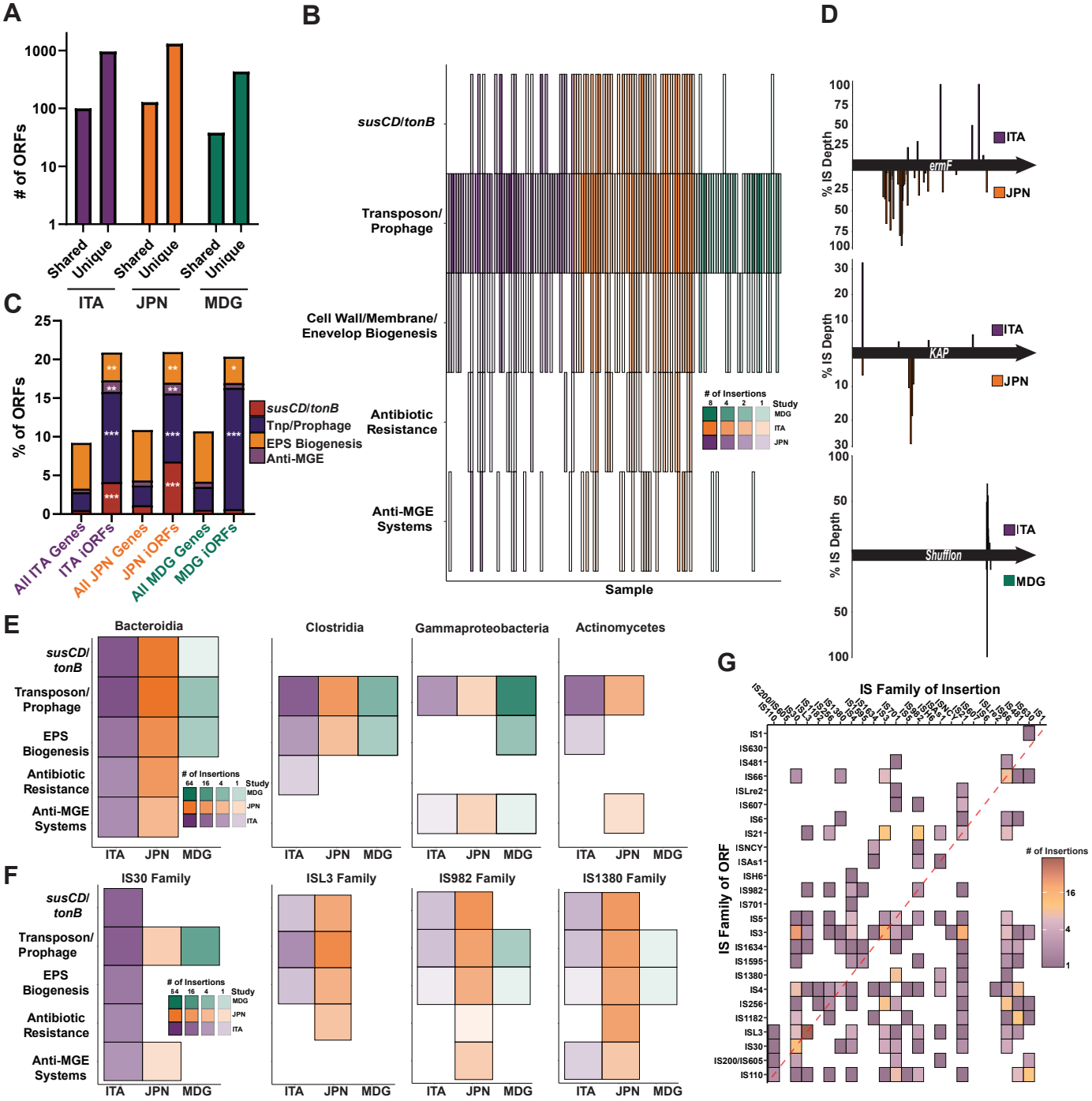
114. Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res 28:33-6. 10.1093/nar/28.1.33

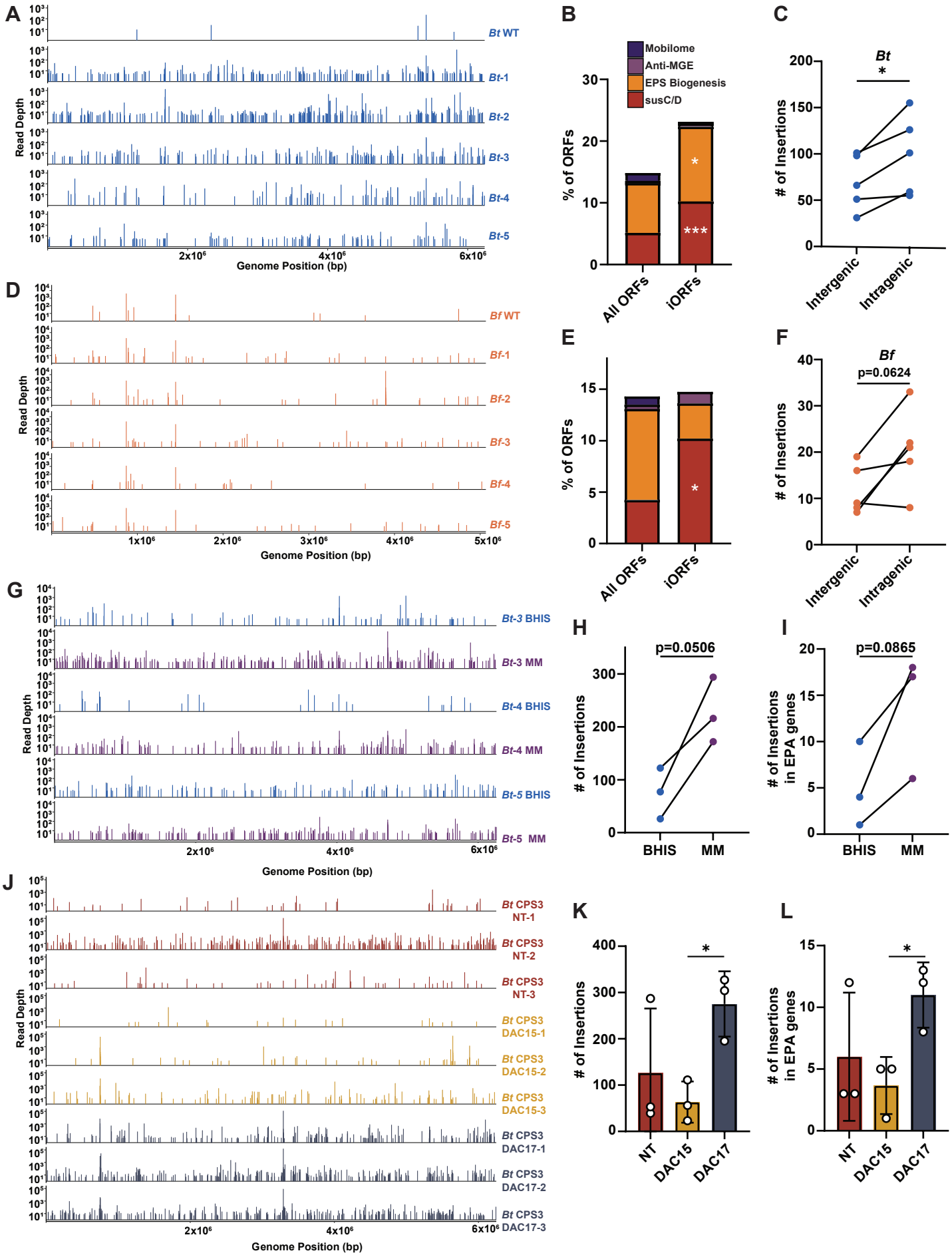
115. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen AV, Cheng AA, Liu S, Min SY, Miroshnichenko A, Tran HK, Werfalli RE, Nasir JA, Oloni M, Speicher DJ, Florescu A, Singh B, Faltyn M, Hernandez-Koutoucheva A, Sharma AN, Bordeleau E, Pawlowski AC, Zubyk HL, Dooley D, Griffiths E, Maguire F, Winsor GL, Beiko RG, Brinkman FSL, Hsiao WWL, Domselaar GV, McArthur AG. 2020. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res 48:D517-d525. 10.1093/nar/gkz935

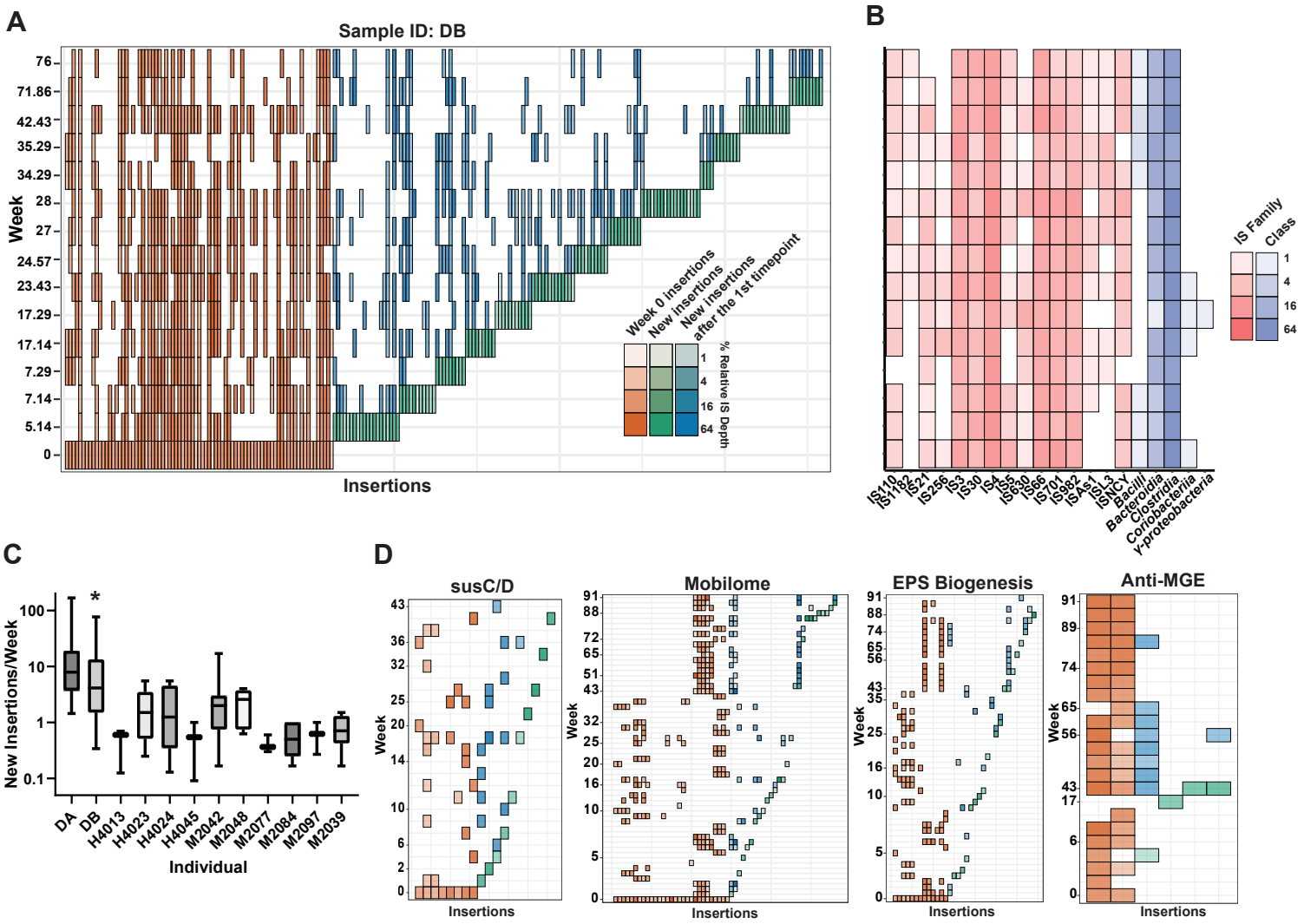
116. Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. Genome Biol 20:257. 10.1186/s13059-019-1891-0

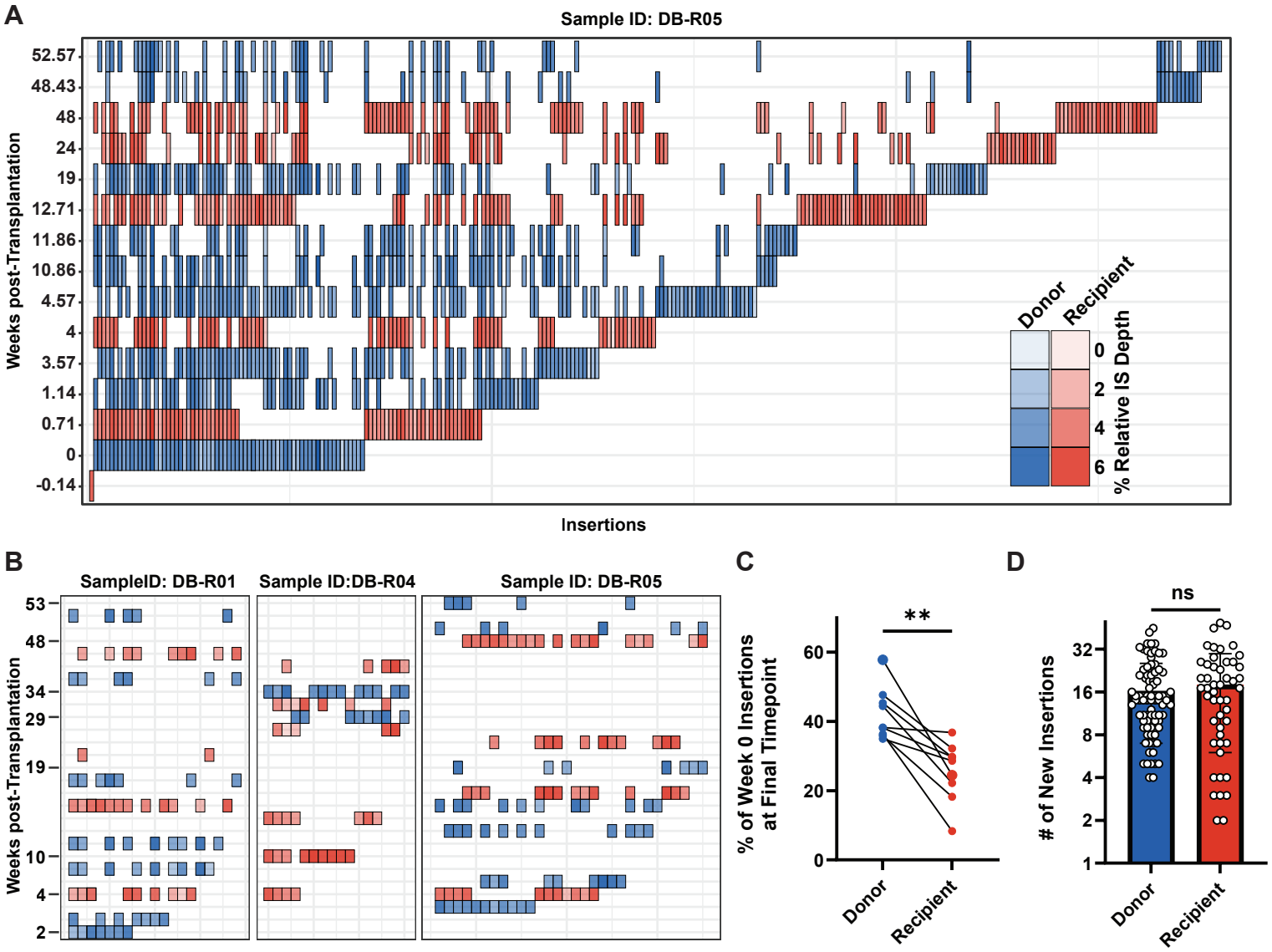

A
# of IS Elements
25000
20000
15000
10000
5000
0
ISFinder
ISOSDB
B
# of IS Elements
20000
15000
10000
5000
0
Tnp
Amino Acid
IS Element
Nucleotide
Homologs in ISFinder
Novel
C
Tree Scale: 10
IS3
IS4
IS256
IS5
IS1182
IS481
IS21
IS630

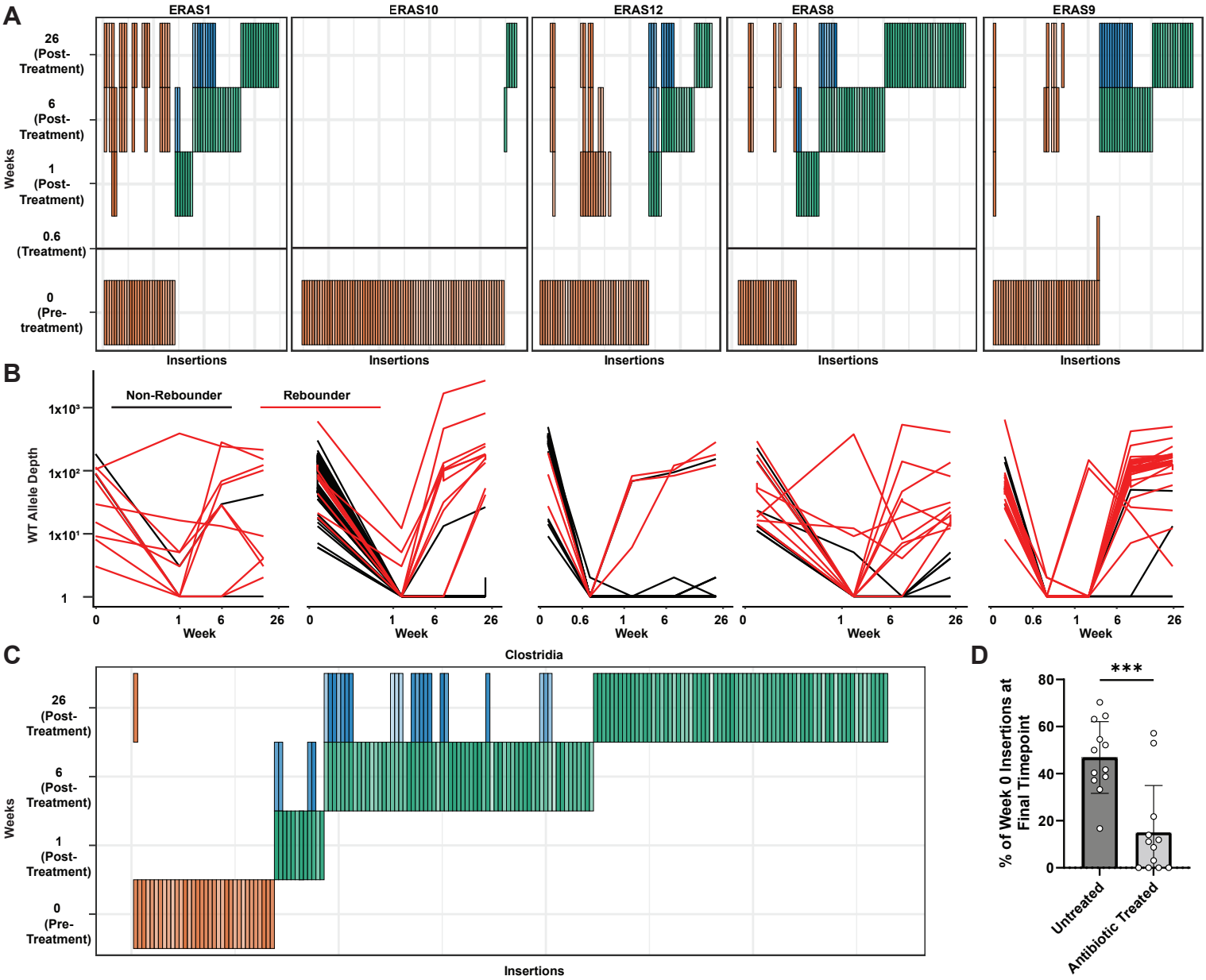

A
ITA
JPN
MDG
100
10
1
% Relative IS Depth
Sample
B
100
10
1
# of Insertions
ITA
JPN
MDG
C
Actinomycetes
Bacilli
Bacteroidia
Clostridia
Coriobacteriia
Gammaproteobacteria
Class
Sample
# of Insertions
64 16 4 1
Study
MDG
JPN
ITA
D
Bacteroidia
ISL3
ISOSDB4584
ISOSDB6187
ISOSDB12026
ISAs1
ISOSDB178
IS982
ISOSDB634
IS66
ISOSDB250
ISOSDB644
ISOSDB445
ISOSDB2678
IS630
ISOSDB426
IS30
ISOSDB412
IS256
ISOSDB45
IS1380
ISOSDB157
ISOSDB18121
IS1182
ISOSDB33
IS Element | IS Family
Sample
Clostridia
ISNCY
ISOSDB531
ISOSDB642
IS982
ISOSDB402
IS701
ISOSDB510
IS66
ISOSDB5997
ISOSDB134
IS630
ISOSDB351
ISOSDB4536
IS5
ISOSDB7176
ISOSDB829
IS4
ISOSDB476
ISOSDB291
ISOSDB237
IS30
ISOSDB238
IS3
ISOSDB805
ISOSDB791
IS21
ISOSDB556
IS Element | IS Family
Sample
E
ISOSDB18121
ISOSDB1159
ISOSDB237
ISOSDB445
ISOSDB805
ISOSDB412
ISOSDB178
-log10 (Adjusted p Value)
3
2
1
-7.5
-5.0
-2.5
0.0
2.5
5.0
Average (Intragenic - Intergenic)

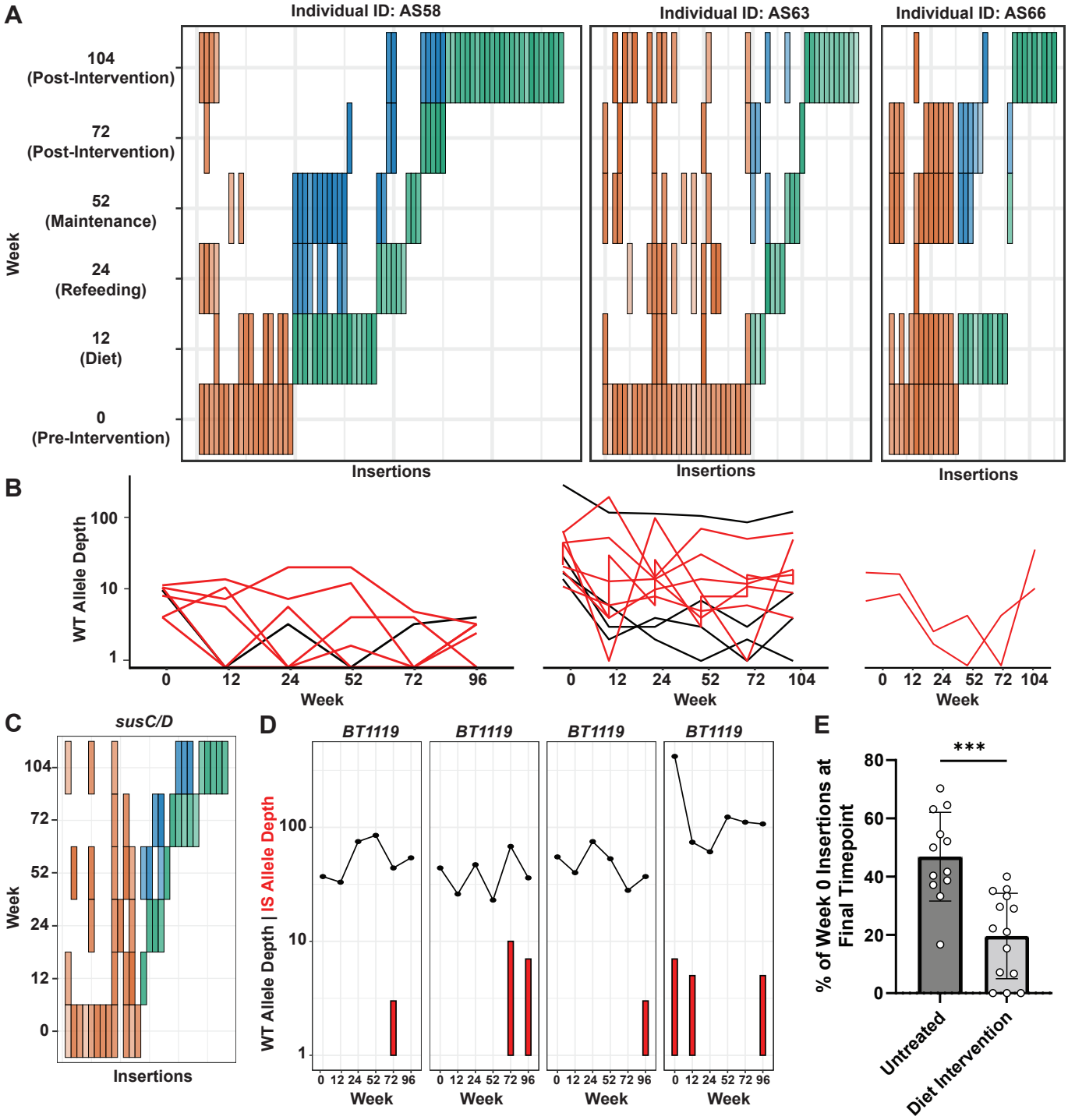

A
# of ORFs
1000
100
10
1
Shared
Unique
ITA
JPN
MDG
B
susCD/tonB
Transposon/ Prophage
Cell Wall/Membrane/ Enevelop Biogenesis
Antibiotic Resistance
Anti-MGE Systems
Sample
# of Insertions
Study
MDG
ITA
JPN
C
% of ORFs
susCD/tonB
Tnp/Prophage
EPS Biogenesis
Anti-MGE
All ITA Genes
ITA iORFs
All JPN Genes
JPN iORFs
All MDG Genes
MDG iORFs
D
% IS Depth
ermF
KAP
Shufflon
ITA
JPN
MDG
E
Bacteroidia
Clostridia
Gammaproteobacteria
Actinomycetes
susCD/ tonB
Transposon/ Prophage
EPS Biogenesis
Antibiotic Resistance
Anti-MGE Systems
ITA
JPN
MDG
F
IS30 Family
ISL3 Family
IS982 Family
IS1380 Family
G
IS Family of Insertion
IS Family of ORF
# of Insertions


A
Bt WT
Bt-1
Bt-2
Bt-3
Bt-4
Bt-5
Read Depth
2x10⁶
4x10⁶
6x10⁶
Genome Position (bp)
B
Mobilome
Anti-MGE
EPS Biogenesis
susC/D
% of ORFs
All ORFs
iORFs
C
Bt
*
# of Insertions
Intergenic
Intragenic
D
Bf WT
Bf-1
Bf-2
Bf-3
Bf-4
Bf-5
Read Depth
1x10⁶
2x10⁶
3x10⁶
4x10⁶
5x10⁶
Genome Position (bp)
E
% of ORFs
All ORFs
iORFs
F
Bf
p=0.0624
# of Insertions
Intergenic
Intragenic
G
Bt-3 BHIS
Bt-3 MM
Bt-4 BHIS
Bt-4 MM
Bt-5 BHIS
Bt-5 MM
Read Depth
2x10⁶
4x10⁶
6x10⁶
Genome Position (bp)
H
p=0.0506
# of Insertions
BHIS
MM
I
p=0.0865
# of Insertions in EPA genes
BHIS
MM
J
Bt CPS3 NT-1
Bt CPS3 NT-2
Bt CPS3 NT-3
Bt CPS3 DAC15-1
Bt CPS3 DAC15-2
Bt CPS3 DAC15-3
Bt CPS3 DAC17-1
Bt CPS3 DAC17-2
Bt CPS3 DAC17-3
Read Depth
2x10⁶
4x10⁶
6x10⁶
Genome Position (bp)
K
*
# of Insertions
NT
DAC15
DAC17
L
*
# of Insertions in EPA genes
NT
DAC15
DAC17


A
Sample ID: DB
Week
Insertions
Week 0 insertions
New insertions
New insertions after the 1st timepoint
% Relative IS Depth
B
IS Family
Class
IS110
IS1182
IS21
IS256
IS3
IS30
IS4
IS5
IS630
IS66
IS701
IS982
ISAs1
ISL3
ISNCY
Bacilli
Bacteroidia
Clostridia
Coriobacteria
γ-proteobacteria
C
New Insertions/Week
Individual
DA
DB
H4013
H4023
H4024
H4045
M2042
M2048
M2077
M2084
M2097
M2039
D
susC/D
Mobilome
EPS Biogenesis
Anti-MGE
Week
Insertions


A
Sample ID: DB-R05
Weeks post-Transplantation
52.57
48.43
48
24
19
12.71
11.86
10.86
4.57
4
3.57
1.14
0.71
0
-0.14
Insertions
Donor
Recipient
% Relative IS Depth
0
2
4
6
B
SampleID: DB-R01
Sample ID:DB-R04
Sample ID: DB-R05
Weeks post-Transplantation
53
48
34
29
19
10
4
2
C
% of Week 0 Insertions at Final Timepoint
**
60
40
20
0
Donor
Recipient
D
ns
# of New Insertions
32
16
8
4
2
1
Donor
Recipient


A
ERAS1
ERAS10
ERAS12
ERAS8
ERAS9
Weeks
26 (Post-Treatment)
6 (Post-Treatment)
1 (Post-Treatment)
0.6 (Treatment)
0 (Pre-treatment)
Insertions
B
Non-Rebounder
Rebounder
WT Allele Depth
1x10³
1x10²
1x10¹
1
Week
C
Clostridia
Weeks
26 (Post-Treatment)
6 (Post-Treatment)
1 (Post-Treatment)
0 (Pre-Treatment)
Insertions
D
% of Week 0 Insertions at Final Timepoint
***
Untreated
Antibiotic Treated


A
Individual ID: AS58
Individual ID: AS63
Individual ID: AS66
104 (Post-Intervention)
72 (Post-Intervention)
52 (Maintenance)
24 (Refeeding)
12 (Diet)
0 (Pre-Intervention)
Week
Insertions
B
WT Allele Depth
100
10
1
0 12 24 52 72 96
0 12 24 52 72 104
Week
C
susC/D
104
72
52
24
12
0
Week
Insertions
D
BT1119
WT Allele Depth | IS Allele Depth
100
10
1
0 12 24 52 72 96
Week
E
***
% of Week 0 Insertions at Final Timepoint
80
60
40
20
0
Untreated
Diet Intervention