

In vitro demonstration of broad host range mobile genetic elements transferring antibiotic resistance from the human microbiome

Samuel C. Forster^{1,2,3,*}, Junyan Liu¹, Nitin Kumar¹, Emily L. Gulliver^{2,3}, Jodee A. Gould^{2,3}, Alejandra Escobar-Zepeda¹, Tapoka Mkandawire¹, Lindsay J. Pike¹, Yan Shao¹, Mark D. Stares¹, Hilary P. Browne¹, B. Anne Neville¹ and Trevor D. Lawley^{1,*}

¹Host-Microbiota Interactions Laboratory, Wellcome Sanger Institute, Hinxton, CB10 1SA, UK,

²Centre for Innate Immunity and Infectious Diseases, Hudson Institute of Medical Research, Clayton, Victoria, 3168, Australia,

³Department of Molecular and Translational Sciences, Monash University, Clayton, Victoria, 3800, Australia

* corresponding authors

Samuel C. Forster: Wellcome Sanger Institute, Hinxton, Cambridgeshire, UK, CB10 1SA, Phone 01223 495 391, Fax 01223 495 239, Email: sf15@sanger.ac.uk

Trevor D. Lawley: Wellcome Sanger Institute, Hinxton, Cambridgeshire, UK, CB10 1SA, Phone 01223 495 391, Fax 01223 495 239, Email: tl2@sanger.ac.uk

Key words: gastrointestinal microbiota, antibiotic resistance, AMR, microbiome, plasmids, mobile genetic elements

Mobile genetic elements (MGEs) carrying antibiotic resistance genes (ARGs) disseminate ARGs when they mobilise into new bacterial hosts. To investigate the scale of such horizontal gene transfer (HGT) events between human gut commensals and enteropathogens for the first time, we compared 1354 cultured commensal strains (540 species) to 45,403 enteropathogen strains (12 species) and found 64,188 MGE-mediated ARG transfer events between the two groups using established methods. Among the 5931 MGEs involved, we found 15 broad host range elements predicted to have crossed different bacterial phyla while also occurring in animal and environmental microbiomes. We experimentally demonstrated that predicted broad host range MGEs can mobilise from commensals *Dorea longicatena* and *Hungatella hathewayi* to enteropathogens *Klebsiella oxytoca*, crossing phyla simultaneously. Our work establishes the MGE-mediated ARG dissemination network between human gut commensals and enteropathogens and highlights novel broad host range MGEs as targets for future ARG dissemination management.

Humans are colonized by microbial communities dominated by bacteria from the Firmicutes, Bacteroidetes, Actinobacteria and Proteobacteria phyla that play an essential role regulating immune response¹, aiding sustenance², and providing pathogen colonization resistance³. Antibiotic treatment, though intended to eliminate pathogens, simultaneously eradicates indigenous commensal bacteria that are sensitive to the antibiotic. This can result in a microbiome with a vastly altered community structure and function; however, commensal species with intrinsic or acquired antibiotic resistance are protected from elimination. Antibiotic selection likely results in antibiotic resistance genes (ARGs) accumulation among commensal bacteria⁴ that may also act as a reservoir from which ARGs are transferred on mobile genetic elements (MGEs) to other species, including pathogens, via horizontal gene transfer (HGT)⁵⁻¹¹.

The extent of HGT in human gut microbiome and the types of MGEs involved, especially those associated with ARGs, have been the focus of continued research interest in the last decade. Several key metagenomic studies showed that HGT between gut bacteria is more frequent than HGT with bacteria from other body sites or environments because intestinal bacteria occupy the same habitat⁵. Most MGEs involved mobilise within the same phylum or lower taxonomic groups⁸ and the transfer of ARGs between pathogens and commensals is considered limited^{5, 12}. A recent bioinformatic study utilising two separate genome collections has demonstrated a capacity to accurately predict compatible HGT host-recipient pairs¹¹. However, experimental validation of HGT has largely relied on high-throughput chromatin conformation capture (Hi-C) which has uncovered extensive HGT in situ including between pathogenic and commensal species^{9, 10}. Many studies using animal models have also demonstrated individual cases of increased HGT between pathogens and commensals during infections^{13, 14}. Despite these advances, the majority of large-scale studies still do not experimentally validate MGE mobility at the isolate level.

It is therefore imperative to understand the true scale of HGT between pathogenic and commensal species coexisting in the human gut, to identify MGEs posing high risks in order to better inform future interventions to curb spread of ARGs. We and others have recently demonstrated that the vast majority of the human gastrointestinal bacteria can be cultured¹⁵⁻²⁰. We used this resource to systematically investigate the extent of HGT between pathogens and commensals, with a focus on ARG-associated MGEs. We were able to confirm host range of MGEs with high confidence, strain-level resolution and validate in vitro putative past HGT events to demonstrate that these MGEs retain the ability to mobilise and can potentially spread associated ARG to numerous bacteria species.

To map the extent of horizontally shared MGEs that carry ARGs between pathogenic and commensal bacteria of the human gastrointestinal tract, we first compared 1354 commensal genomes (530 species) from the Human Gastrointestinal Microbiota Genome Collection (HGG)²¹ (Supplementary Table 1) to 45,403 publicly available genomes of human enteropathogenic and opportunistic pathogenic species (Supplementary Table 2). These genomes represented 12 gastrointestinal pathogenic species including 8 Proteobacteria; *Klebsiella oxytoca* (n=139), *K. pneumoniae* (n=7712), *Escherichia coli* (n=17142), *Salmonella enterica* (n=10394), *Shigella sonnei* (n=1290), *S. flexneri* (n=453), *Campylobacter coli* (n=919) and *C. jejuni* (n=1554) and 4 Firmicutes; *Enterococcus faecalis* (n=1364), *E. faecium* (n=1706), *Clostridioides difficile* (n=2016) and *Clostridium perfringens* (n=138) (Fig. 1a). Some of the 12 species are among the most prevalent gastrointestinal pathogens globally²²⁻²⁴, some are posing great risk to the public as they become increasingly resistant to antibiotics²⁵.

We reasoned that genes originating either directly or indirectly through recent horizontal transfer events would exhibit high nucleotide homology between isolates incongruent with phylogenetic distance. Pairwise gene comparisons were performed between

the 1354 commensal genomes and 45,403 enteropathogen genomes to identify those genes sharing significant nucleotide identity (>99% identity across over 500bp in organisms <97% 16S rRNA homology⁵) between strains from either group. This analysis found 389,541 putative horizontally transferred genes (Fig. 1c).

To further identify genes encoding antibiotic resistance within the 389,541 putatively transferred genes, we computationally defined ARGs by comparison to the Comprehensive Antibiotic Resistance Database (CARD)²⁶. This identified 64,188 (16.5%) of the putative horizontally acquired genes shared between commensals and pathogens to be ARGs. These ARGs were dominated by multidrug efflux complexes, aminoglycoside resistance, cationic antimicrobial peptide resistance and beta-lactamases (Supplementary Fig. 1). Notably, we observed no statistically significant enrichment in ARG class or gene families in either pathogen or commensal genomes. While these results reflect that pathogen associated genes dominate ARG databases^{26, 27}, they also suggest no obvious barriers to ARG dissemination within either pathogens or commensals and between the two groups although we cannot determine the directionality of these HGT events. These observed patterns suggest that ARG dissemination networks are highly interconnected within the gut microbiome⁵ and further demonstrate that they transcend the type of bacterial symbiosis with humans.

To define the MGEs responsible for mediating ARG transfer events between pathogens and commensals, we next consolidated the 64,188 shared ARGs into the common genetic elements by combining elements with greater than 90% shared homology across the element. A total of 5931 MGEs were identified within the dataset through this process. Analysis of these elements demonstrated a range of ARGs including dihydrofolate reductases, tetracycline resistance and aminoglycoside resistance with no enrichment for any individual class relative to the occurrence within the larger dataset.

The pathogens *E. faecalis* (n=1364 genomes), *C. difficile* (n=2016 genomes) and *E. faecium* (n=1706 genomes) from the Firmicutes phylum shared the greatest number of distinct genetic elements with the commensal isolates with significant enrichment by Fisher exact test relative to numbers of genomes ($q < 10^{-6}$, $q < 10^{-4}$ and $q < 10^{-6}$). In contrast, the pathogens from the Proteobacteria phylum including, *Escherichia coli* ($q < 10^{-12}$; n=17142 genomes), *K. pneumoniae* ($q < 10^{-6}$; n=7712 genomes) and *Shigella sonnei* ($q < 10^{-3}$; n=1290 genomes), were significantly under-represented by Fisher Exact Test when evaluating the overall number of shared distinct MGEs with the commensal microbiota. Equally, commensal Bacteroidetes and Actinobacteria shared few ARG containing MGEs with any of the pathogenic species considered in this study. While no events were observed with Fusobacteria, the small number of genomes limited statistical interpretation. Thus, although ARG-associated MGEs are widespread, MGE diversity and abundances are not evenly distributed across gut bacterial phyla, with Firmicutes exhibiting the highest enrichment of MGE diversity ($p < 0.05$; Fisher Exact Test) and sharing the most MGEs between commensals and pathogens ($p < 0.05$; Fisher Exact Test).

Our ability to confidently assign specific MGEs to bacterial host species provided a unique opportunity to generate a phylogenetic framework to study MGE host range. It is known that the horizontal transfer range of a MGE is limited by phylogenetic barriers of the bacterial hosts⁸, although this has not been investigated at large scale across human gut bacteria. We found that 79.33% of MGEs transferred within a genus, 11% transferred across classes and, interestingly, 1.5% of MGEs had a broad host range being found across multiple phyla (Fig. 1b). Our results demonstrate the widespread presence of phylogenetic barriers which prevent the majority of MGEs disseminating broadly across bacterial phyla, with the majority of ARG transfer being restricted to within a bacterial genus.

Given the obvious risks to human health associated with the capacity to disperse between bacterial phyla, we focused our analysis on those MGEs capable of broad host range and wide-scale dissemination between pathogens and commensals inferred from diverse phylogenetic occurrence. This analysis identified 15 broad host range MGEs shared between more than one bacterial phylum within our dataset. To provide a greater understanding of the genomic architecture surrounding the identified elements we coupled long-read sequencing technology with automated and manual annotations to generate high-quality reference genomes for the 15 broad host range MGEs (Supplementary Information).

The 15 promiscuous MGEs include three plasmids (PLASMID_1-3), three integrative and conjugative elements (ICEs; ICE_1-3) and nine integrative and mobilizable elements (IMEs; IME_1-9) (Supplementary Table 3). While the majority (n=14) of these elements could be found in commensal Firmicutes, six were also found in commensal Actinobacteria, five in commensal Bacteroidetes and three in commensal Proteobacteria (Fig. 2). Elements occurred in a median of 5 of the 12 pathogenic species within this study (41.6%; min: 3; max: 10) with elements in three Firmicutes (75%; min=1, max=4) and two Proteobacteria (25%; min=0, max=7) species. Between one and four ARGs were found on each element, with tetracycline and aminoglycosides being the most common encoded resistance. There were no common genetic elements that provide a universal explanation for the observed broad host range.

We next classified the broad host range MGEs in the context of previously reported incompatibility (Inc) groups²⁸⁻³⁰ based on the type of MGE replication machinery. We examined each element for the presence of a gene encoding RepA protein using PSI-BLAST (cutoff E-value<1e-05). Our analysis identified putative RepA proteins in eight MGEs that share limited similarity (between 20% and 40%) with known Inc groups, and one element, PLASMID_1 (Supplementary Fig. 2), that contains a putative RepA which is not closely

related to any known Inc group (Supplementary Table 4). Thus, nine MGEs have putative novel replication proteins, but no identifiable genes encoding RepA could be recognized in the remaining six elements (Supplementary Table 3). We found that four elements (IME_1, IME_5, ICE_3, IME_8) encode RepA proteins that belong to a clade including the RepA found in IncQ plasmids and one element (PLASMID_2) that encodes a RepA that belongs to a clade that contains RepA found in IncA/C plasmids (Fig. 3a). In another cluster, RepA proteins from another three elements (ICE_1, ICE_2, IME_4) share ancestry with RepA from IncP1 plasmids (Fig. 3b). Importantly, MGEs belonging to IncA/C, IncQ and IncP1 groups commonly carry ARGs in Proteobacteria^{19,20} but MGEs with related replication machineries have not been previously recognized as drivers of broad host range ARG dissemination in Firmicutes and Bacteroidetes from the human microbiome.

Despite the observed wide host range of these ARG-associated MGEs based on our genomic analysis, the HGT events could be ancient. We next used in vitro conjugation assay to test the ability of these MGEs, especially ICEs and plasmids, to mobilise between commensals and pathogens of different classes or phyla. ICE_1 carries *aadK* (streptomycin resistance) and can be found in a *Faecalicatena contorta* strain (Clostridia class of Firmicutes). ICE_2 carries *tetM* (tetracycline resistance) and can be found in a *Hungatella hathewayi* strain (Clostridia class of Firmicutes). Both ICEs can be transferred into a streptomycin and tetracycline sensitive *E. faecalis* strain (Bacilli class of Firmicutes) by dry patch conjugation, demonstrating their ability to mobilise across different classes (Fig. 3c&d; Supplementary Fig. 3). Based on earlier analysis of this study, ICE_2 is predicted to be able to mobilise across phyla as well. We further selected a novel plasmid PLASMID_1 which carries a poorly characterised tetracycline resistant determinant and can be found in a *Dorea longicatena* strain (Lachnospiraceae family of Firmicutes). Using filter and dry patch conjugation respectively, ICE_2 and PLASMID_1 were transferred into a tetracycline

sensitive *Klebsiella oxytoca* strain (Enterobacteriaceae family of Proteobacteria), demonstrating the ICE and plasmid still retain the ability to mobilise between different bacterial phyla (Fig. 3e; Supplementary Fig. 3). In all four pairs of conjugation assays, transfer of the MGE of interest into recipient bacterial strains was confirmed by PCR on all transconjugants (Supplementary Fig. 4) and long read sequencing of some randomly selected transconjugants. Interestingly, we noted that formerly tetracycline sensitive *E. faecalis* strain, after receiving either PLASMID_1 or ICE_2, exhibit higher MIC (minimal inhibitory concentration) than the original donor strains (Fig. 3d&e). This implies the same ARG containing MGE can be responsible for quantitatively variable antibiotic resistant phenotypes in different bacterial host species within the gastrointestinal microbiota.

To understand the prevalence and environmental range of the 15 broad host range elements we next examined 4,349 high coverage human gastrointestinal microbiome-associated metagenomes available within the HPMC database³¹. We determined prevalence rates of these elements to be between 0.52% and 98.2% with no statistically significant enrichment observed with Fisher Exact Test based on element type, size or antibiotic resistance genes carried (Fig. 4). To understand if these elements or associated homologues are also found within the microbiome communities of other human body sites, we determined the occurrence and prevalence of these elements within the skin, nasal cavity and female reproductive tract datasets of the Human Microbiome Project dataset. This analysis identified five elements limited in distribution to the human gastrointestinal tract, eight were detected in samples from the human nasal cavity, nine were detected in samples from the human vagina and three were detected in samples from the human skin with a coverage of 99% and detection rate of 0.001% (Fig. 4). Considered together, these results suggest an interconnection within the human microbiome, potentially limited by movement of species between body sites³². Equivalent analysis of samples from ruminant gastrointestinal

microbiome identified the presence of six elements, all of which can be also detected in both human gut, nasal cavity and vagina. On the other hand, only three elements were detected within environmental soil samples (Fig. 4). Hence, we demonstrate the presence of identical broad range MGEs in humans, animals and environmental sources, highlighting the need for a One Health approach to understand ARG distribution.

While the prevalence of ARGs within human gut microbiome has been well established, we are lacking an understanding of ARG-associated MGE host range and prevalence in the context of the entire human microbiome. Our work represents the first large-scale, whole-genome, strain-level analysis coupled with experimental validation to generate an MGE-mediated ARG dissemination network between human commensals and enteropathogens. This work characterizes the diversity and host range of MGEs harboured by the gut microbiome, demonstrates the retained ability of key, novel broad host range MGEs to mobilize between diverse commensal and pathogenic species. In a medical context, these novel broad host range MGEs may represent a significant threat aiding ARG dispersal independently of the infection control measures established to contain specific pathogens. These insights suggest effective antimicrobial stewardship will require a focus not only on controlling antibiotic resistant pathogens but tracking, managing and limiting ARG-associated MGE dissemination³³ from both pathogenic and commensal bacterial species.

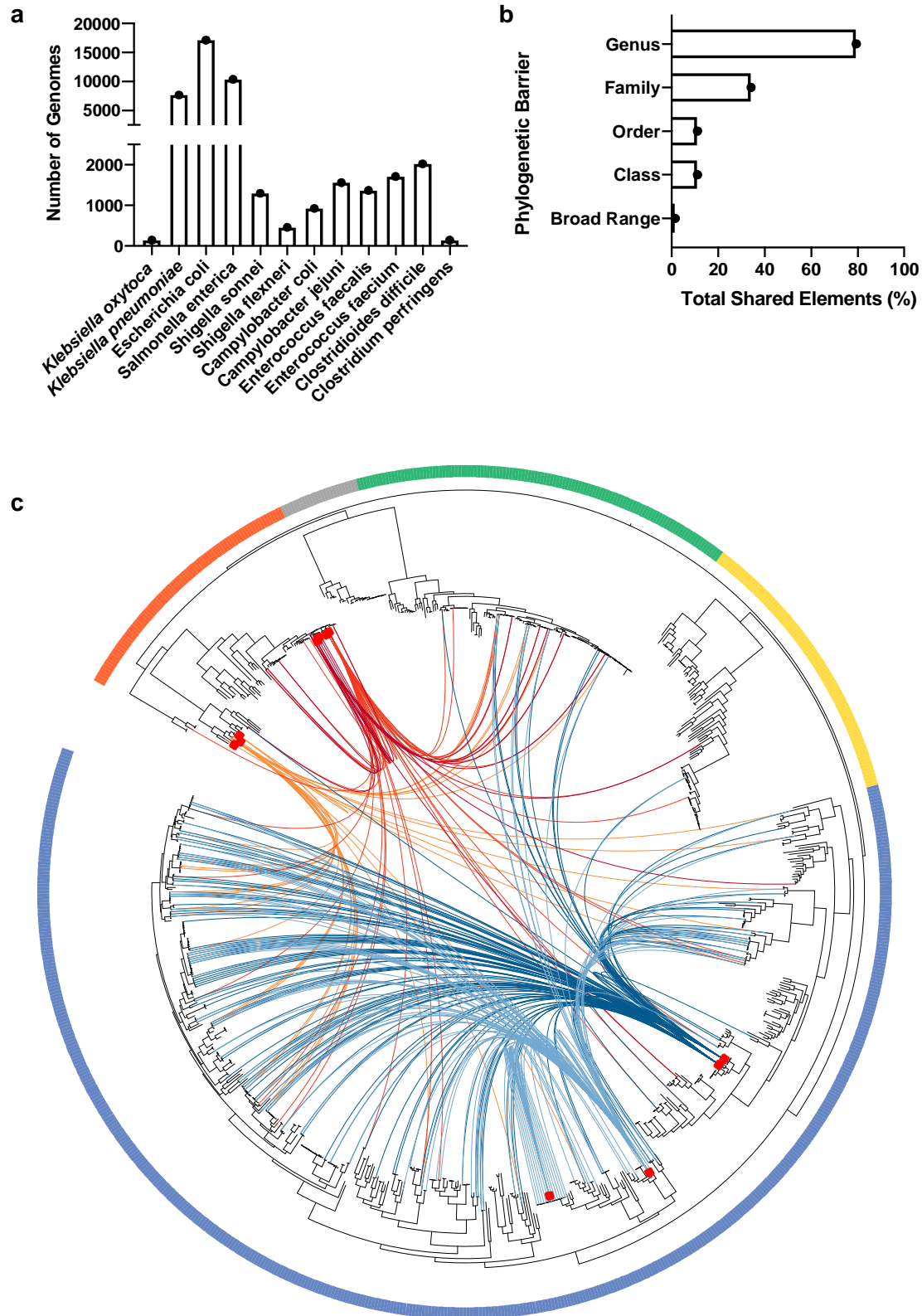


Figure 1: Patterns of horizontally transferred genetic elements between gastrointestinal commensal and pathogen species.

a, Number of genomes from each of the twelve pathogenic species included in the analysis.

b, Phylogenetic barriers of MGE demonstrating intra-genus (79.3%), intra-family (34.2%), intra-order (11.14%), intra-class (11.12%) and broad host range, intra-phyla (1.5%) horizontal transfer. **c**, Actinobacteria (Gold), Bacteroidetes (Green), Firmicutes (Blue), Fusobacteria (Grey) and Proteobacteria (Orange) are indicated by the surrounding border. Red boxes indicate location of pathogens on the phylogenetic tree. Connecting lines indicate shared putative horizontally transferred genes. Links from *Escherichia* (Red), *Klebsiella* (Dark Red), *Campylobacter* (Orange), *Enterococcus* (Dark Blue) and *Clostridia* (Light Blue) pathogens are shown.

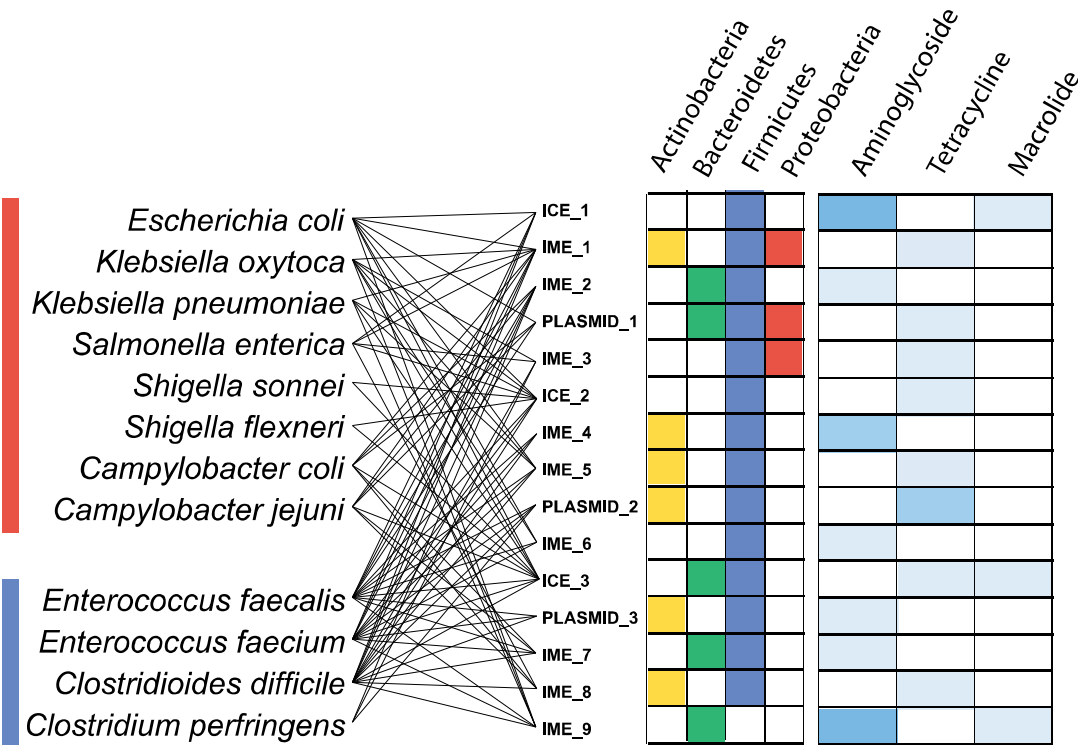


Figure 2: Taxonomic distribution and functional composition of 15 broad host range mobile genetic elements.

Occurrence of 15 promiscuous elements shared between pathogenic bacteria and commensal Actinobacteria (yellow), Bacteroidetes (green), Firmicutes (blue) and Proteobacteria (red). The presence of 1 (light blue), 2 (moderate blue) or 3 (dark blue) ARGs (Aminoglycoside, Tetracycline, Macrolide). Connecting lines show the presence of promiscuous elements in respective pathogenic bacteria.

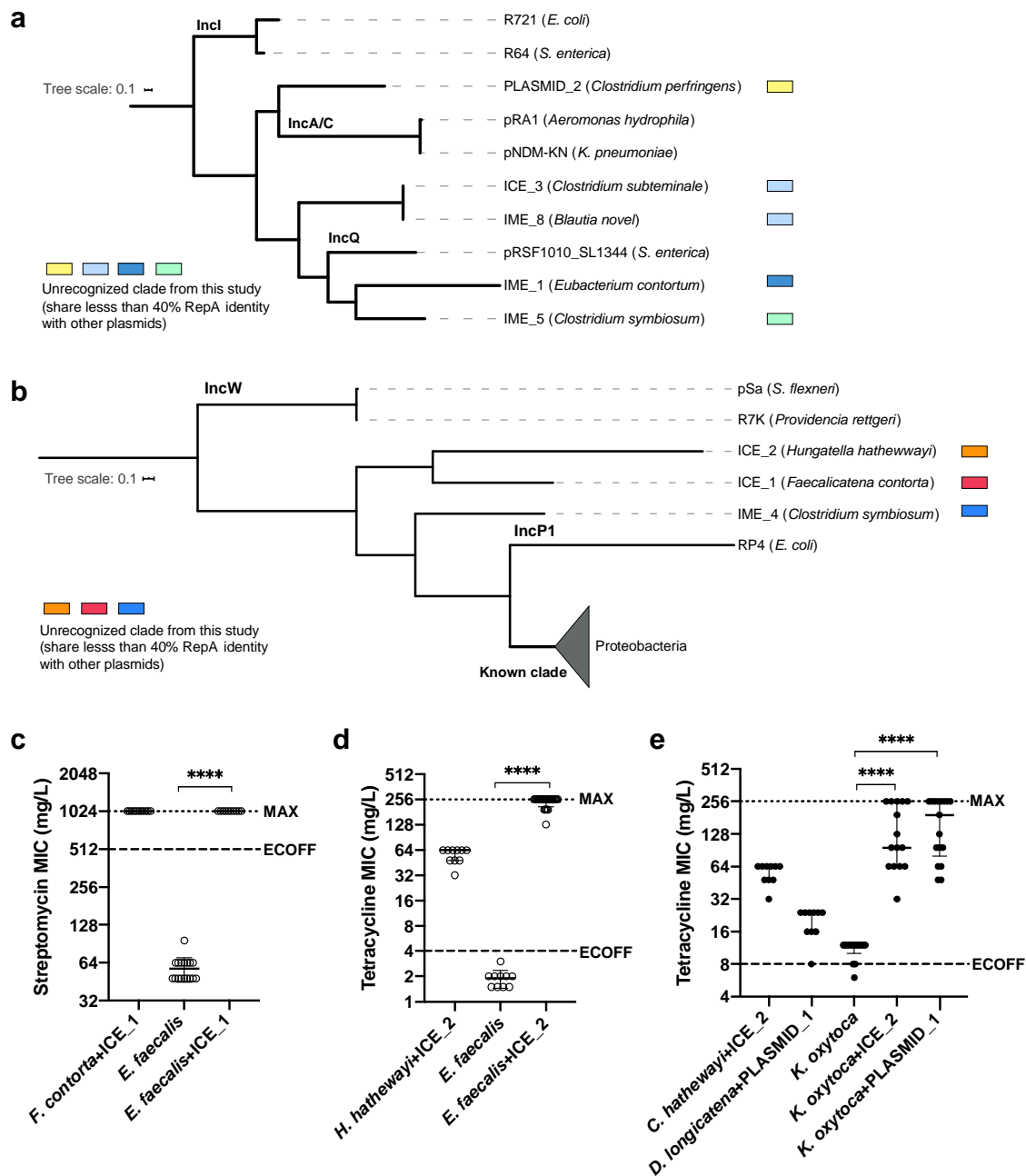


Figure 3: Classification of novel, broad host range mobile genetic elements and experimental validation of horizontal transfer between distantly related species.

a, b, Phylogeny of the RepA proteins from 8 elements with representatives from closely related incompatibility groups. **c,** ICE_1 from *F. contorta* and **d,** ICE_2 from *H. hathewayi* can conjugate into *E. faecalis*. **e,** ICE_2 and plasmid PLASMID_1 from *D. longicatena* can be transferred into *K. oxytoca* by conjugations. The median and interquartile range of each

data set is indicated. The dashed line indicates the epidemiological cut-off values (ECOFF) for minimal inhibitory concentration (MIC) of either streptomycin or tetracycline against the recipient (**b**, $n = 11, 17, 10$; **d**, $n = 10, 10, 30$; **e**, $n = 10, 10, 13, 15, 17$; two-tailed Mann-Whitney test, **** P -value < 0.0001), while the dotted line indicates the maximal antibiotic concentration of respective E-strips (MAX).

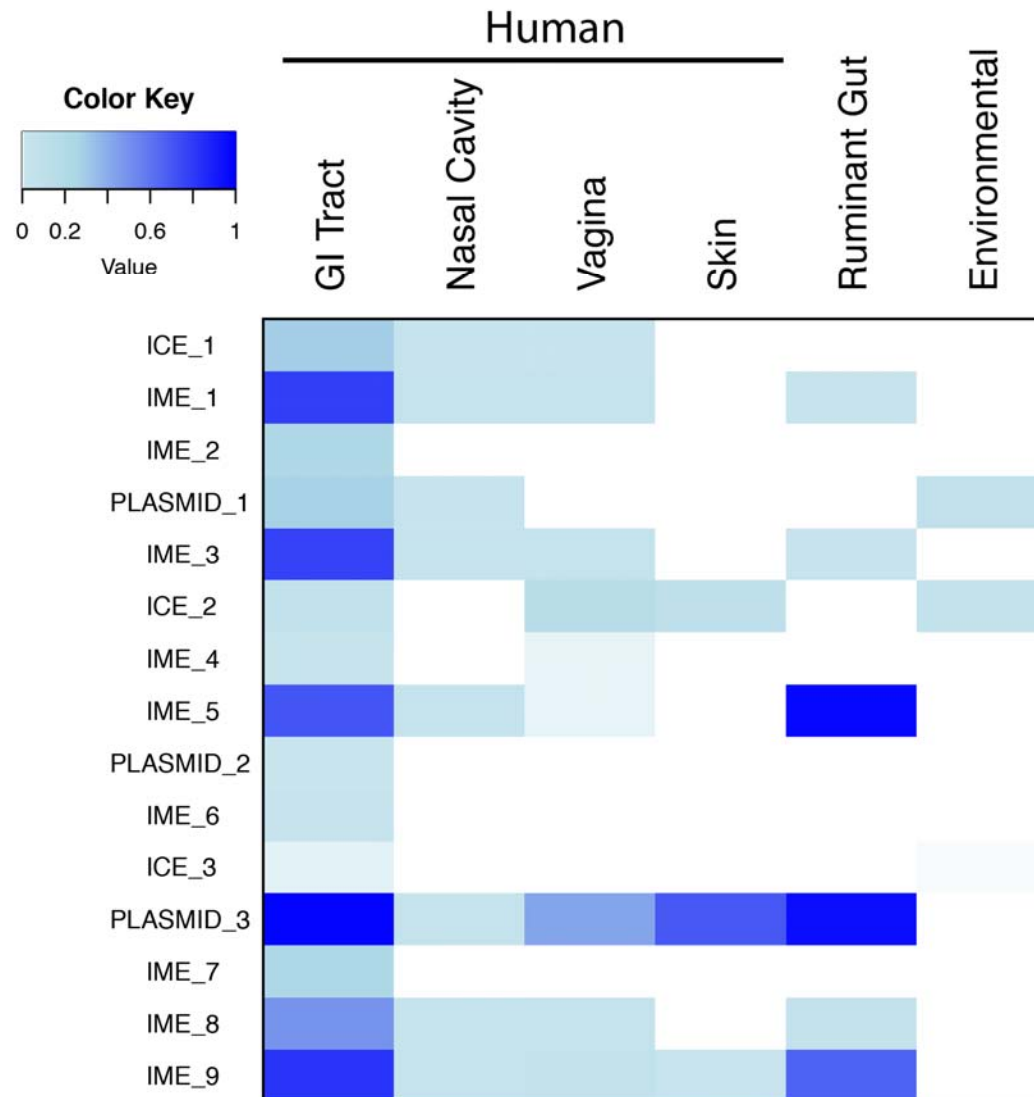


Figure 4: Prevalence of the promiscuous mobile elements in human body, ruminant gut and environmental microbiomes.

Prevalence of promiscuous elements across the human gastrointestinal tract, nasal cavity, vagina, skin, ruminant gut and environmental samples. Blue represents the frequency with which an element was detected within a metagenomic sample derived from the site. No detection is represented in white.

Methods

Bacterial culturing

Culturing was performed from fresh faecal samples under anaerobic conditions in a Whitley A95 anaerobic workstation (Don Whitley) as described previously¹⁵. Standard culturing was performed on YCFA media at 37°C and colonies were picked 72 hours after plating and re-streaked to confirm purity. Selective culturing was performed anaerobically and aerobically on YCFA media supplemented with antibiotic at indicated concentration at 37°C.

Bacterial conjugation and antibiotic sensitivity testing

Donor and recipient isolates were grown in YCFA broth anaerobically overnight without shaking. For filter conjugation, donor and recipient cultures were diluted to OD₆₀₀=1, then mixed in 4:1 ratio and immobilized onto a sterile 0.45 µm mixed cellulose ester membrane (Whatman) which was then placed on a YCFA plate with the side containing bacterial cells facing up. For dry patch conjugation, donor and recipient cultures were concentrated to OD₆₀₀=40 and OD₆₀₀=10 respectively, mixed by equal volume (4:1 ratio by cell density) and spotted on YCFA plate. All conjugation plates were incubated anaerobically overnight and washed with 1-1.5 mL PBS to harvest cells. The cell suspension was plated on YCFA plates with or without antibiotic for transconjugants selection and enumeration respectively. The concentration of streptomycin used is 300 µg/mL. The concentrations of tetracycline used is 50 µg/mL (low selection) and 60 µg/mL (high selection). These plates were incubated aerobically to select against obligatory anaerobic donors. All the recipient controls were treated following the same protocol respectively. The MIC of specific antibiotic against donor, recipient and transconjugants were measured using E-strips (bioMerieux) according to manufacturer's instructions on YCFA media. Donor isolates for ICE_1 (DSM 108236; JCM 31258), PLASMID_1 (DSM 108233) and ICE_2 (CCUG 68736) are available in public depository.

Polymerase chain reaction (PCR) for mobile element detection

PCR was performed to confirm transfer of mobile element into recipient bacteria. ICE_1 forward primer (TGATATCATGGAAGGTCGGCA) and reverse primer (ACCGCCCTGAACAATTGATG) target a 181 bp region within the *aadK* gene. PLASMID_1 forward primer (AAAGCAGCTATCATTCGGGT) and reverse primer (TGCCCGCCTTTGAAGATACC) target a 301 bp region within the *mobC* gene. ICE_2 forward primer (TTGATGCCCTTTTGAAATC) and reverse primer (ACTGCATTCCACTTCCCAAC) target a 294 bp region within *tetM* gene.

Statistical tests

Two-tailed Mann-Whitney test (unpaired, nonparametric) was used to compare MIC or conjugation frequency data where applicable.

Short read genome sequencing and annotation

Genomic DNA was extracted from bacterial cell pellets using a phenol-chloroform method described previously¹⁵. Libraries were prepared and paired end- sequencing performed on the Illumina Hi-Seq platform with a target library fragment size of 450 bp and a read length of 150 bp at the Wellcome Sanger Institute according to standard protocols. Annotated assemblies were produced using the pipeline standard Wellcome Sanger Institute prokaryotic assembly and annotation pipeline³⁴. Briefly, for each sample an optimal assembly was created using Velvet v1.2³⁵ and VelvetOptimiser v2.2.5, scaffolded using SSPACE v3.0³⁶ and sequence gaps filled using GapFiller v1.10³⁷. PROKKA v1.14³⁸ was used for automated gene annotation. Genome sequences have been deposited in the European Nucleotide Archive.

Long read genome sequencing and annotation

Genomic DNA was extracted from bacterial cell pellets using MasterPure™ complete DNA

purification kit (Lucigen). Oxford Nanopore Minion Sequencing, multiplexed libraries were sequenced using the ligation sequencing kit according to manufacturer's instructions. Where required, genomes were demultiplexed using guppy v1.1 and assembled using canu v1.9³⁹. Within genomes mobile genetic elements were located by the ARG previously identified. The element boundaries were determined as a consensus from predictions by IslandViewer 4⁴⁰, ICEfinder v1.0⁴¹, Alien_Hunter v1.7⁴² and the presence of inverted and direct repeat sequences.

Phylogenetics and Reference Genomes

The phylogenetic analysis was conducted by extracting amino acid sequence of 40 universal single copy marker genes^{43, 44} from bacterial collection using SpecI v1.0⁴⁵. The protein sequences were concatenated and aligned with MAFFT v. 7.20⁴⁶, and maximum-likelihood trees were constructed using raxML v8.2.11 with default settings. All phylogenetic trees were visualized in iTOL v5.6⁴⁷. Average nucleotide analysis (ANI) was calculated by performing pairwise comparison of genome assemblies using MUMmer v3.0⁴⁸. Genomes from 12 pathogenic species were obtained from NCBI based on taxonomic classification.

Horizontally Acquired Gene Analysis

Shared genes were identified by pairwise BLAST v2.6.0+ of annotated genes (greater than 99% identity across 500bp or greater sequence). Putative horizontally acquired genes were identified by shared gene in organisms with greater than 97% 16S rRNA homology⁷. The CARD database version 1.9 was used to identify ARG with a 95% identity and 90% coverage cutoff.

Mobile Element Identification and Novelty

Complete mobile element sequences were retrieved from the PLSDb plasmid database v. 2019_06_03⁴⁹ and the ICEberg 2.0 database. Mash v2.1.1 sketches were generated (-i -S 42, -

k 21 -s 1000) and putative mobile elements screened with Mash screen (-v 0.1, -i 0.95). 950 of 1000 shared Mash hashes was used as the distance cutoff for novel element identification.

Replication Initiation Protein (RepA) Classification

RepA proteins were identified by PSI-BLAST v2.6.0+ (Position-Specific Iterated BLAST) against a plasmid replication protein database created from three studies⁵⁰⁻⁵² with a cutoff E-value of 1e-05 and RepA protein sequences available on NCBI.

Metagenomic Prevalence

Prevalence of mobile elements was assessed using bowtie2 v2.3.4.1 searches of metagenomic datasets retrieved from the HPMC database²¹ and the HMP database. Elements were considered present within the sample when occurring at greater than 0.001% of sample with greater than 99% coverage.

References

1. Rooks MG, Garrett WS. Gut microbiota, metabolites and host immunity. *Nat Rev Immunol.* May 27 2016;16(6):341-52. doi:10.1038/nri.2016.42
2. LeBlanc JG, Milani C, de Giori GS, Sesma F, van Sinderen D, Ventura M. Bacteria as vitamin suppliers to their host: a gut microbiota perspective. *Current Opinion in Biotechnology.* Apr 2013;24(2):160-168. doi:10.1016/j.copbio.2012.08.005
3. Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. Diversity, stability and resilience of the human gut microbiota. *Nature.* Sep 13 2012;489(7415):220-30. doi:10.1038/nature11550
4. Jakobsson HE, Jernberg C, Andersson AF, Sjolund-Karlsson M, Jansson JK, Engstrand L. Short-Term Antibiotic Treatment Has Differing Long-Term Impacts on the Human Throat and Gut Microbiome. *Plos One.* Mar 24 2010;5(3)doi:ARTN e9836 10.1371/journal.pone.0009836
5. Smillie CS, Smith MB, Friedman J, Cordero OX, David LA, Alm EJ. Ecology drives a global network of gene exchange connecting the human microbiome. *Nature.* Dec 8 2011;480(7376):241-244. doi:10.1038/nature10571
6. Hu Y, Yang X, Qin J, et al. Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut microbiota. *Nat Commun.* 2013;4:2151. doi:10.1038/ncomms3151
7. Hu Y, Yang X, Li J, et al. The Bacterial Mobile Resistome Transfer Network Connecting the Animal and Human Microbiomes. *Appl Environ Microbiol.* Nov 15 2016;82(22):6672-6681. doi:10.1128/AEM.01802-16
8. Jiang X, Hall AB, Xavier RJ, Alm EJ. Comprehensive analysis of chromosomal mobile genetic elements in the gut microbiome reveals phylum-level niche-adaptive gene pools. *PLoS One.* 2019;14(12):e0223680. doi:10.1371/journal.pone.0223680
9. Kent AG, Vill AC, Shi QJ, Satlin MJ, Brito IL. Widespread transfer of mobile antibiotic resistance genes within individual gut microbiomes revealed through bacterial Hi-C. *Nature Communications.* Sep 1 2020;11(1)doi:ARTN 4379 10.1038/s41467-020-18164-7
10. Yaffe E, Relman DA. Tracking microbial evolution in the human gut using Hi-C reveals extensive horizontal gene transfer, persistence and adaptation. *Nature Microbiology.* Feb 2020;5(2):343-+. doi:10.1038/s41564-019-0625-0
11. Ellabaan MMH, Munck C, Porse A, Imamovic L, Sommer MOA. Forecasting the dissemination of antibiotic resistance genes across bacterial genomes. *Nat Commun.* Apr 23 2021;12(1):2435. doi:10.1038/s41467-021-22757-1
12. Ruppe E, Ghazlane A, Tap J, et al. Prediction of the intestinal resistome by a three-dimensional structure-based method. *Nat Microbiol.* Jan 2019;4(1):112-123. doi:10.1038/s41564-018-0292-6
13. Zhang XP, McDaniel AD, Wolf LE, Keusch GT, Waldor MK, Acheson DWK. Quinolone antibiotics induce shiga toxin-encoding bacteriophages, toxin production, and death in mice. *Journal of Infectious Diseases.* Feb 2000;181(2):664-670. doi:10.1086/315239
14. Stecher B, Denzler R, Maier L, et al. Gut inflammation can boost horizontal gene transfer between pathogenic and commensal Enterobacteriaceae. *Proceedings of the National Academy of Sciences of the United States of America.* Jan 24 2012;109(4):1269-1274. doi:10.1073/pnas.1113246109
15. Browne HP, Forster SC, Anonye BO, et al. Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. *Nature.* 05 2016;533(7604):543-6. doi:10.1038/nature17645

16. Goodman AL, Kallstrom G, Faith JJ, et al. Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. *Proc Natl Acad Sci U S A*. Apr 12 2011;108(15):6252-7. doi:10.1073/pnas.1102938108
17. Lagier JC, Khelaifia S, Alou MT, et al. Culture of previously uncultured members of the human gut microbiota by culturomics. *Nat Microbiol*. Nov 7 2016;1:16203. doi:10.1038/nmicrobiol.2016.203
18. Lau JT, Whelan FJ, Herath I, et al. Capturing the diversity of the human gut microbiota through culture-enriched molecular profiling. *Genome Medicine*. Jul 1 2016;8doi:ARTN 72
10.1186/s13073-016-0327-7
19. Poyet M, Groussin M, Gibbons SM, et al. A library of human gut bacterial isolates paired with longitudinal multiomics data enables mechanistic microbiome research. *Nat Med*. Sep 2019;25(9):1442-1452. doi:10.1038/s41591-019-0559-3
20. Zou Y, Xue W, Luo G, et al. 1,520 reference genomes from cultivated human gut bacteria enable functional microbiome analyses. *Nat Biotechnol*. Feb 2019;37(2):179-185. doi:10.1038/s41587-018-0008-8
21. Forster SC, Kumar N, Anonye BO, et al. A human gut bacterial genome and culture collection for improved metagenomic analyses. *Nat Biotechnol*. Feb 2019;37(2):186-192. doi:10.1038/s41587-018-0009-7
22. Fletcher SM, McLaws ML, Ellis JT. Prevalence of gastrointestinal pathogens in developed and developing countries: systematic review and meta-analysis. *J Public Health Res*. Apr 28 2013;2(1):42-53. doi:10.4081/jphr.2013.e9
23. Balsells E, Shi T, Leese C, et al. Global burden of Clostridium difficile infections: a systematic review and meta-analysis. *J Glob Health*. Jun 2019;9(1):010407. doi:10.7189/jogh.09.010407
24. Lindstrom M, Heikinheimo A, Lahti P, Korkeala H. Novel insights into the epidemiology of Clostridium perfringens type A food poisoning. *Food Microbiology*. Apr 2011;28(2):192-198. doi:10.1016/j.fm.2010.03.020
25. Asokan GV, Ramadhan T, Ahmed E, Sanad H. WHO Global Priority Pathogens List: A Bibliometric Analysis of Medline-PubMed for Knowledge Mobilization to Infection Prevention and Control Practices in Bahrain. *Oman Med J*. May 2019;34(3):184-193. doi:10.5001/omj.2019.37
26. Jia B, Raphenya AR, Alcock B, et al. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res*. Jan 4 2017;45(D1):D566-D573. doi:10.1093/nar/gkw1004
27. Yin X, Jiang XT, Chai B, et al. ARGs-OAP v2.0 with an expanded SARG database and Hidden Markov Models for enhancement characterization and quantification of antibiotic resistance genes in environmental metagenomes. *Bioinformatics*. Jul 1 2018;34(13):2263-2270. doi:10.1093/bioinformatics/bty053
28. Carattoli A. Resistance Plasmid Families in Enterobacteriaceae. *Antimicrobial Agents and Chemotherapy*. Jun 2009;53(6):2227-2238. doi:10.1128/Aac.01707-08
29. Harmer CJ, Hall RM. The A to Z of A/C plasmids. *Plasmid*. Jul 2015;80:63-82. doi:10.1016/j.plasmid.2015.04.003
30. Popowska M, Krawczyk-Balska A. Broad-host-range IncP-1 plasmids and their resistance potential. *Frontiers in Microbiology*. Mar 7 2013;4doi:ARTN 44
10.3389/fmicb.2013.00044
31. Forster SC, Browne HP, Kumar N, et al. HPMCD: the database of human microbial communities from metagenomic datasets and microbial reference genomes. *Nucleic Acids Res*. Jan 4 2016;44(D1):D604-9. doi:10.1093/nar/gkv1216

32. Thomas-White K, Forster SC, Kumar N, et al. Culturing of female bladder bacteria reveals an interconnected urogenital microbiota. *Nature Communications*. Apr 19 2018;9doi:ARTN 1557
10.1038/s41467-018-03968-5
33. Cabezon E, de la Cruz F, Arechaga I. Conjugation Inhibitors and Their Potential Use to Prevent Dissemination of Antibiotic Resistance Genes in Bacteria. *Frontiers in Microbiology*. Nov 30 2017;8doi:ARTN 2329
10.3389/fmicb.2017.02329
34. Page AJ, De Silva N, Hunt M, et al. Robust high-throughput prokaryote de novo assembly and improvement pipeline for Illumina data. *Microb Genom*. Aug 2016;2(8):e000083. doi:10.1099/mgen.0.000083
35. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res*. May 2008;18(5):821-9. doi:10.1101/gr.074492.107
gr.074492.107 [pii]
36. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics*. Feb 15 2011;27(4):578-579. doi:10.1093/bioinformatics/btq683
37. Boetzer M, Pirovano W. Toward almost closed genomes with GapFiller. *Genome Biology*. 2012;13(6)doi:ARTN R56
10.1186/gb-2012-13-6-r56
38. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. Jul 15 2014;30(14):2068-9. doi:10.1093/bioinformatics/btu153
btu153 [pii]
39. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res*. May 2017;27(5):722-736. doi:10.1101/gr.215087.116
40. Bertelli C, Laird MR, Williams KP, et al. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res*. Jul 3 2017;45(W1):W30-W35. doi:10.1093/nar/gkx343
41. Liu M, Li XB, Xie YZ, et al. ICEberg 2.0: an updated database of bacterial integrative and conjugative elements. *Nucleic Acids Research*. Jan 8 2019;47(D1):D660-D665. doi:10.1093/nar/gky1123
42. Vernikos GS, Parkhill J. Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the Salmonella pathogenicity islands. *Bioinformatics*. Sep 15 2006;22(18):2196-2203. doi:10.1093/bioinformatics/btl369
43. Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P. Toward automatic reconstruction of a highly resolved tree of life. *Science*. Mar 3 2006;311(5765):1283-1287. doi:10.1126/science.1123061
44. Sorek R, Zhu YW, Creevey CJ, Francino MP, Bork P, Rubin EM. Genome-wide experimental determination of barriers to horizontal gene transfer. *Science*. Nov 30 2007;318(5855):1449-1452. doi:10.1126/science.1147112
45. Mende DR, Sunagawa S, Zeller G, Bork P. Accurate and universal delineation of prokaryotic species. *Nature Methods*. Sep 2013;10(9):881-+. doi:10.1038/Nmeth.2575
46. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*. Apr 2013;30(4):772-780. doi:10.1093/molbev/mst010
47. Letunic I, Bork P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Research*. Jul 2011;39:W475-W478. doi:10.1093/nar/gkr201

48. Delcher AL, Phillippy A, Carlton J, Salzberg SL. Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Research*. Jun 1 2002;30(11):2478-2483. doi:DOI 10.1093/nar/30.11.2478
49. Galata V, Fehlmann T, Backes C, Keller A. PLSDB: a resource of complete bacterial plasmids. *Nucleic Acids Research*. Jan 8 2019;47(D1):D195-D202. doi:10.1093/nar/gky1050
50. Yanagiya K, Maejima Y, Nakata H, et al. Novel Self-Transmissible and Broad-Host-Range Plasmids Exogenously Captured From Anaerobic Granules or Cow Manure. *Front Microbiol*. 2018;9:2602. doi:10.3389/fmicb.2018.02602
51. Yano H, Shintani M, Tomita M, Suzuki H, Oshima T. Reconsidering plasmid maintenance factors for computational plasmid design. *Comput Struct Biotechnol J*. 2019;17:70-81. doi:10.1016/j.csbj.2018.12.001
52. Zhang D, Zhao Y, Feng J, et al. Replicon-Based Typing of IncI-Complex Plasmids, and Comparative Genomics Analysis of IncIgamma/K1 Plasmids. *Front Microbiol*. 2019;10:48. doi:10.3389/fmicb.2019.00048

Acknowledgements

This work was supported by the Wellcome Trust [098051]; the Australian National Health and Medical Research Council [1091097, 1159239 and 1156333 to SCF] and the Victorian Government's Operational Infrastructure Support Program (SCF). The authors would also like to acknowledge the support of the Wellcome Sanger Institute Pathogen Informatics Team and Monash University eResearch.

Funding for open access charge: Wellcome Sanger Institute.

Author Contributions

SCF, BAN, JL, NK and TDL conceived the study. SCF, EG, NK, JAG, TM and AE performed the computational analysis. SCF, BAN, LJP, YS, HPB, JAG and MDS isolated and purified the bacteria and performed genome sequencing. JL, TM performed the experimental analysis. All authors read and contributed to the manuscript.

Data Availability

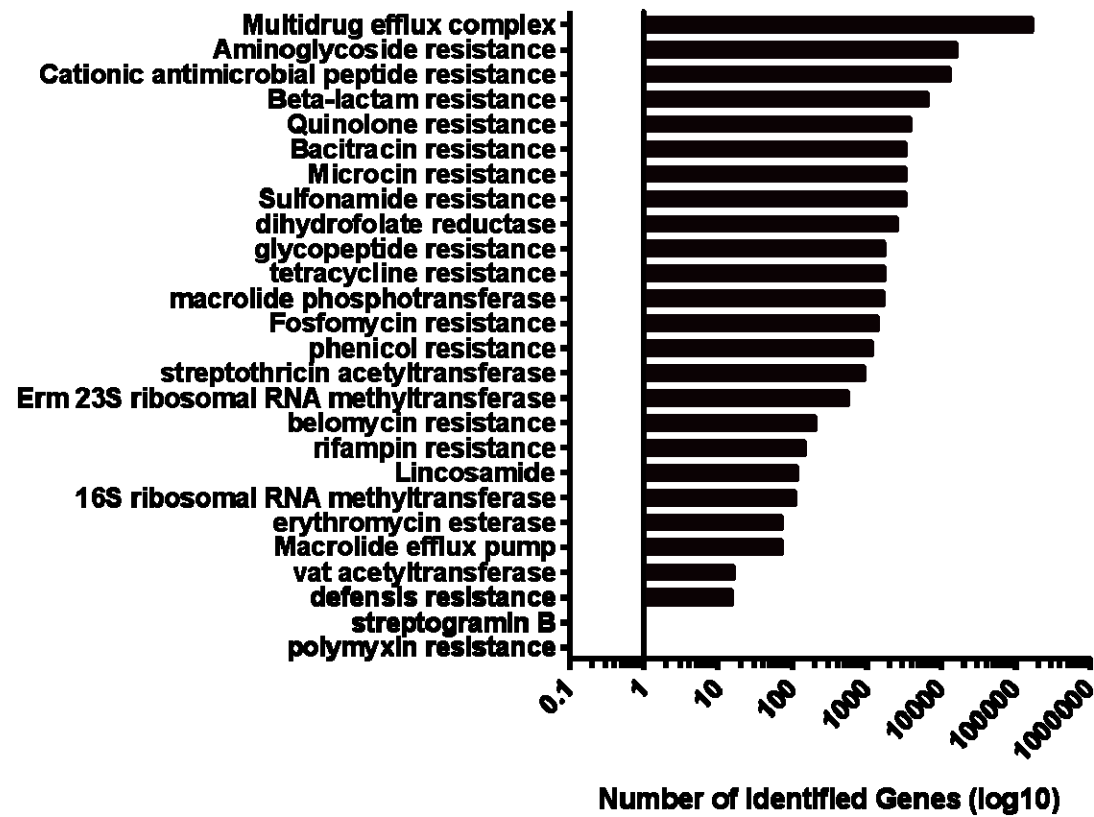
Sequence data is deposited in the ENA under Accession number PRJEB37690. Sequence data from the HGG is deposited in the ENA under project numbers ERP105624 and ERP012217.

Competing Interests

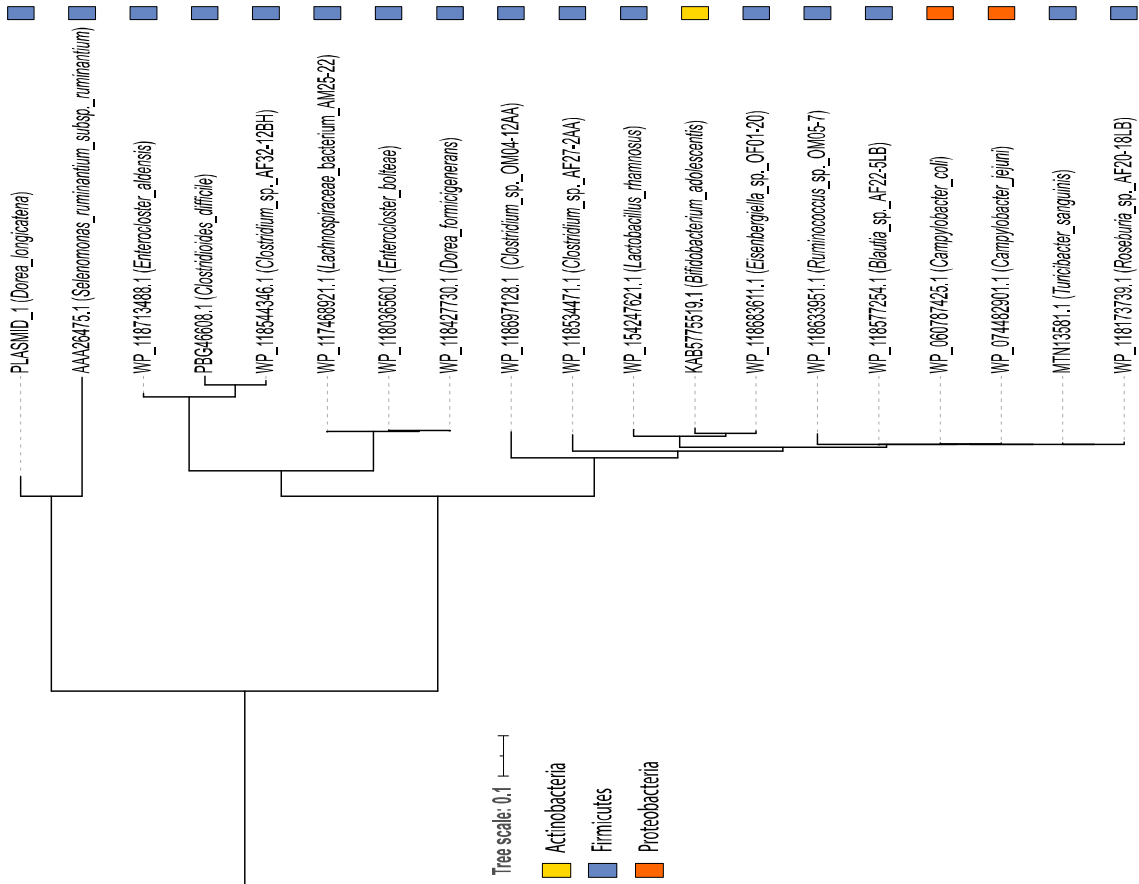
The authors declare no competing financial interests.

Correspondence and requests for materials should be addressed to Trevor D. Lawley or Samuel C Forster

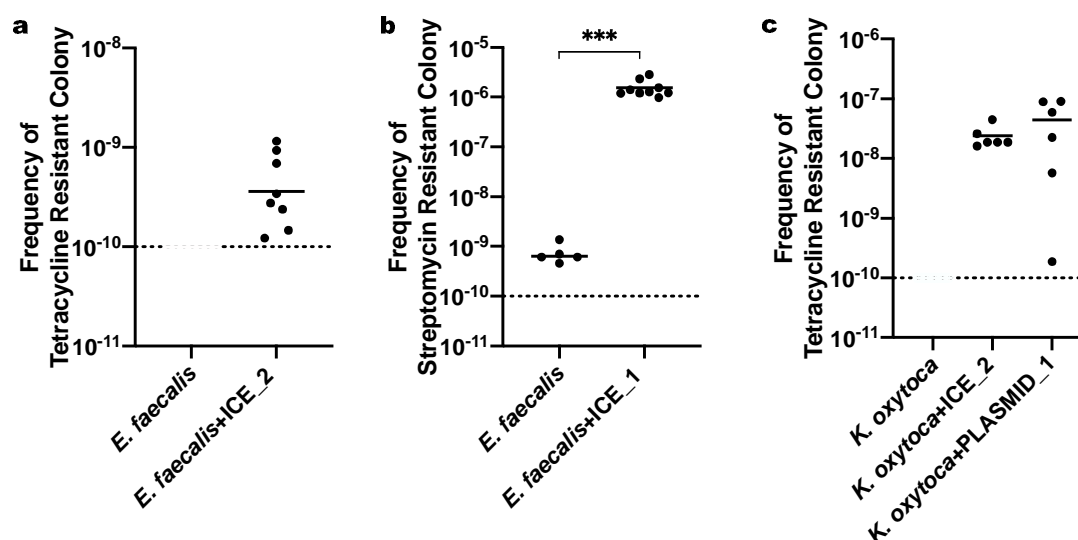
Supplementary Figures



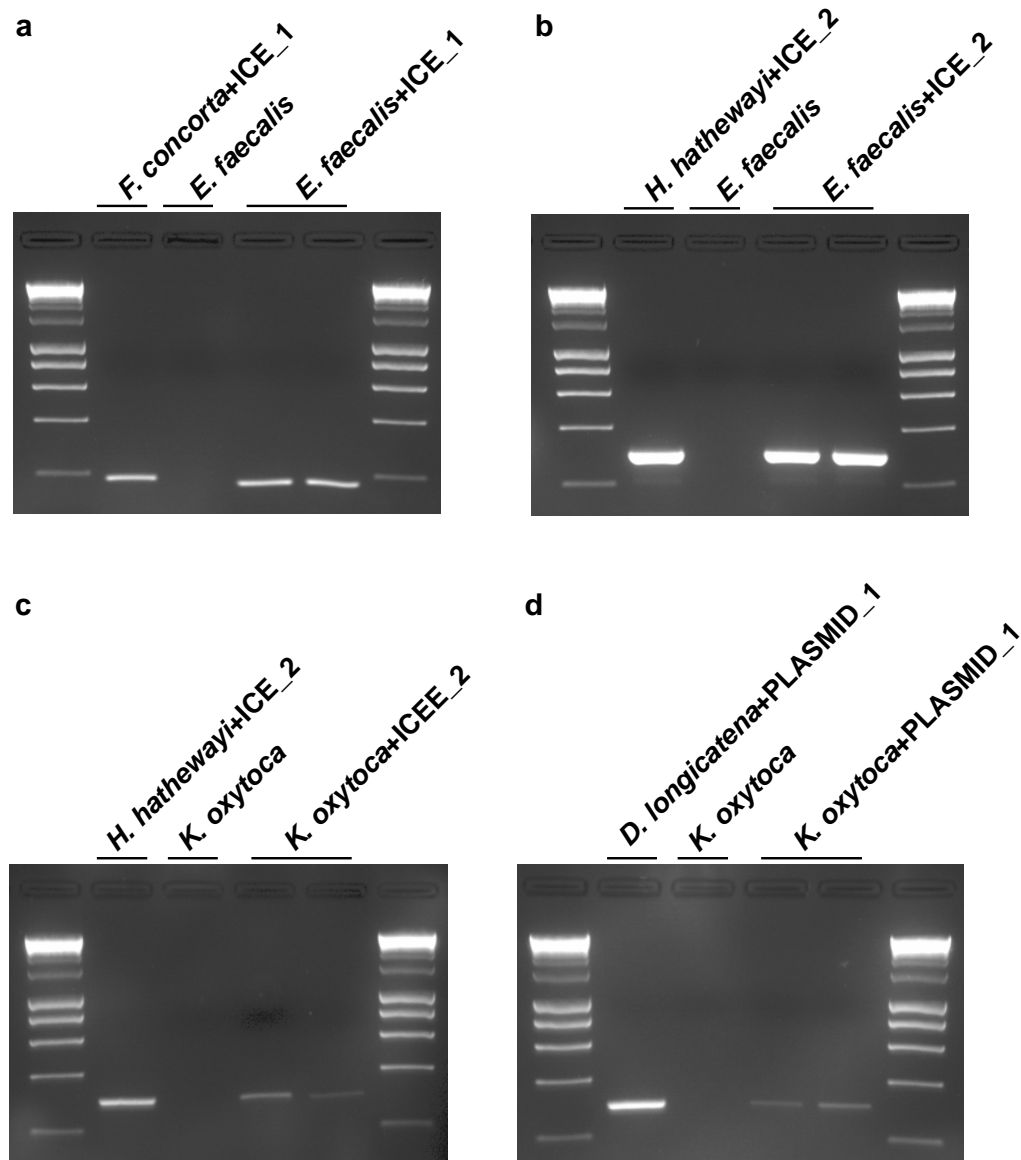
Supplementary Figure 1: ARG classes identified as shared between the pathogenic and symbiont bacterial genomes.



Supplementary Figure 2: Maximum likelihood Tree of RepA from PLASMID_1.
Maximum likelihood tree of RepA from PLASMID_1 and 17 closely related homologues (14 Firmicutes (Blue); 2 Proteobacteria (Red); 1 Actinobacteria (Yellow)).



Supplementary Figure 3: Conjugation frequency of *E. faecalis* and *K. oxytoca* (transconjugants per recipient). **a**, Tetracycline resistance arising from spontaneous mutation was below detection limit 10^{-10} ($n = 12, 12$; high selection). **b**, Streptomycin resistance arisen from spontaneous mutation in the recipient subjected to the same conjugation protocol without the donor was about 1,000 times lower than that of the conjugation (Two-tailed Mann-Whitney test, *** P -value = 0.0004; $n = 6, 9$). **c**, The rate of spontaneous tetracycline resistant mutation was below detection limit 10^{-10} ($n = 12, 6, 6$; low selection). The mean of each data set is indicated. Data points below detection limit are not shown.



Supplementary Figure 4: Agarose gel (2%) electrophoresis of PCR products from all conjugations. a, conjugation of ICE_1 from *F. contorta* into *E. faecalis*; conjugation of ICE_2 from *H. hathewayi* **b**, into *E. faecalis* and **c**, into *K. oxytoca*; **d**, conjugation of PLASMID_1 from *D. longicatena* into *K. oxytoca*. Two transconjugants were selected for each pair of conjugation. HyperLadder™ 1 kb (bioline) used.