Chapter 2: Materials and Methods

All analysis was performed by me, unless otherwise stated.

2.1 Anaerobic gut bacteria culture collection

The human commensal gut microbiota used in this thesis belong to the Human Gastrointestinal Bacteria Culture Collection (HBC)[151] that I helped to develop (published in 2019). The HBC contains 737 gut bacteria isolated from 20 humans, 8 from the UK and 12 from North America[151]. Any 16S rRNA gene sequences of less than 98.7 % similarity to known sequences are considered to belong to novel species, < 94.5 % to novel genera, < 86.5 % to novel families, < 82.0 % novel orders, < 78.5 % novel classes and 75.0 % novel phyla[182].

Genomic DNA was extracted from the 737 HBC isolates using a phenol:chloroform procedure[146,151] by Dr Hilary Browne. DNA was sequenced by the Sanger Sequencing Pipelines (Wellcome Sanger Institute, WSI) team using the Illumina HiSeq platform, generating paired-end reads of 125 or 150 bp; these reads were assembled using the pipeline described by Page *et al.* 2016[146,151,183]. The amino acid sequences of forty core genes were extracted from each genome using FetchMG[184], concatenated and aligned using MAFFT v7.0[185,186]. A phylogeny was inferred using FastTree v2.1.3 SSE3[187,188] and the JTT+CAT model of amino acid evolution.

2.2 Genome-based predictions of antibiotic resistance in the HBC

Raw paired-end sequencing reads for each genome were used as input for the Antibiotic Resistance Identification By Assembly (ARIBA) algorithm[160], which performs local assemblies and maps them against a database of antibiotic resistance genes and mutations (genetic

resistance determinants). In this study, the Comprehensive Antibiotic Resistance Database[163,164] (CARD; version 2.0.2 June 2018) was used with default ARIBA parameters: 90% nucleotide identity for clustering sequences (ARIBA prepareref task), plus minimum 90 % alignment identity, minimum 20 % alignment length, minimum 50 reads assembly coverage and 95 % of gene cluster sequence must be assembled to call a gene cluster present (ARIBA run task). The observed determinants were grouped by the class of antibiotics they are reported to confer resistance to, using the CARD ontology. If a determinant was described as conferring resistance to more than one antibiotic, it was classified as 'nonspecific' antibiotic resistance. The exception is for resistances to Macrolide, Lincosamide, Pleuromutilin and Streptogramin (MLPS) antibiotics, as resistance determinants against these antibiotics can have cross-resistance to each other and are grouped together in a single, separate category. The grouping of these determinants was visualised using Krona[189]. The proportion of identified genetic antibiotic resistance determinants that belonged to a particular antibiotic class were calculated and visualized using Krona[189]. The proportion of isolates with at least one resistance determinant was calculated. The interquartile range (determined using the Tukey method[190]) and mean number of predicted antibiotic resistance determinants in an individual isolate was plotted according to the bacterial phyla of each isolate. This was repeated for the number of antibiotic classes isolates were predicted to be resistant to.

In addition, the presence of predicted resistance to a particular antibiotic class was visualised against a core genome phylogeny (generated as described in section 2.1) using the online interactive Tree of Life (iTOL) tool[191]. To identify which antibiotic resistances were enriched in certain phyla, the proportion of isolates with at least one resistance determinant in each phyla was compared to the proportion of the overall HBC with Fisher exact tests[192] and corrected using two-stage linear step-up procedures of Benjamini, Krieger and Yekutieli[193]. These

analyses were repeated in bacterial families containing more than five HBC isolates and in novel versus characterised isolates, plus the proportion of predicted resistances were determined for each antibiotic in novel versus characterised genomes.

ARIBA was also implemented with the MegaRes[167], ResFinder[170] and SRST2-ARGANNOT[161] databases, plus the CARD-RGI[164] tool was applied, with default parameters to predict antibiotic resistance in the HBC genomes.

2.3 Genome-based predictions of antibiotic resistance in pathogenic genomes

PATRIC[194] was searched for genomes for the ESKAPE pathogens (*E. faecium*, *S. aureus*, *K. pneumoniae*, *A. baumannii*, *P. aeruginosa* and *Enterobacter* species). *E. coli* and *C. difficile* were also included as both are important causes of gut-related bacterial infections, and there are commensal isolates of these species in the HBC. I filtered for bacteria isolated from humans in clinical settings to ensure they were definitely pathogenic isolates. In addition, I chose bacteria isolated from 2010 or after, and from Canada, the US or the UK, to be consistent with the HBC isolates. From this list, I selected isolates that were considered to have 'good' quality genomes and with a high level of completeness (CheckM[195] completeness score equal to or greater than 98 %). PATRIC considers genomes good or poor quality based on summary annotation statistics and from comparisons with other PATRIC genomes after they have been through the PATRIC comprehensive genome analysis service[194]. There were no *E. faecium* isolates with post-2010 dates in the official isolation date meta-data column, but looking at other meta-data columns I identified 37 isolates from 2012 or later, two from 2001, one from 2000 and one from 1997; these were all included in the subsequent analyses. From the resultant list of genomes, I downloaded paired-end short read sequences and

predicted the presence of genetic antibiotic resistance determinants in CARD using ARIBA as in section 2.2. I then repeated the analyses of the proportion of isolates with predicted resistance, the interquartile range and mean number of resistance determinants per isolate, and the proportion of predicted resistances per antibiotic as in section 2.3 for each of the eight pathogenic species and between the pathogenic isolates and HBC isolates.

2.4 Phenotypic antibiotic sensitivity in commensal gut bacteria

HBC isolates were streaked out from glycerol stocks onto modified YCFA[146,151,196] agar (Table 2.1) plates in anaerobic conditions in a Whitley DG250 workstation at 37°C and left for 48 hours. Single colonies were sub-cultured and left to grow for another 48 hours; this was repeated once more, then a single colony for each isolate was used to inoculate 1ml of YCFA broth in a 96 well plate. Three separate inoculations were performed per isolate to allow three biological replicates to be tested. After 48 hours in broth, a cotton swab was dipped in each culture and streaked on a YCFA agar plate three times, turning 60° each time. An Oxoid antibiotic disk dispenser was used to place single-concentration antibiotic disks onto the inoculated agar plates. Mr Mark Stares assisted with these phenotypic tests.

Table 2.1 Modified YCFA media. Volumes are to make 500 ml of YCFA broth or agar. The solutions and mixes are prepared separately and added in the required volume when the media is being prepared. d.H_2O = distilled water.

Ingredient (part number)	Amount		Components of solutions and mixes (part number)	Amount
Before autoclaving			**Resazurin Solution**	
Agar (optional)	8 g		Resazurin (10269990)	0.1 g
Tryptone (84610.0500)	5.0 g		d.H_2O	100 ml
Yeast extract (LP021B)	1.25 g			
$NaHCO_3$ (10583381)	2.0 g		**Mineral Solution I**	
(D)+Glucose (G8270)	1.0 g		K_2HPO_4 (10677623)	3 g
(D)+Maltose (CHE1900)	1.0 g		d.H2O	1 L
(D)+Cellobiose (10207603)	1.0 g			
L-cysteine (30089)	0.5 g		**Mineral Solution II**	
Mineral Solution I	75 ml		KH_2PO^4 (P9791)	3 g
Mineral Solution II	75 ml		$(NH_4)2SO_4$ (A4418)	6 g
Resazurin Solution	0.5 ml		NaCl (10616082)	6 g
Haemin Solution	5 ml		$MgSO_4$ (CHE2458)	0.6 g
Vitamin Solution I	0.5 ml		$CaCl_2$ (dry) (10704492)	0.6 g
d.H_2O	Up to 500 ml		d.H2O	1 L
VFA mix	3.1 ml			
NaOH (CHSO0041)	pH to 7.45		**VFA mix**	
			Acetic acid (10304980)	17 ml
After autoclaving			Propionic acid (15658000)	6 ml
Vitamin Solution II	0.5 ml		n-Valeric acid (10686584)	1 ml
			Iso-valeric acid (129542)	1 ml
			Isobutyric acid (11366766)	1 ml
			Haemin Solution	
			KOH (8143530100)	0.28 g
			Ethanol 95 %	25 ml
			Haemin (10506591)	0.1 g
			d.H_2O	Up to 100 ml
			Vitamin Solution I	
			Biotin (B4501)	5 mg
			Cobalamin (Vitamin B12) (47869)	5 mg
			PABA (4-Aminobenzoic Acid) (A9878)	15 mg
			Folic acid (F7876)	25 mg
			Pyridoxine (P5669)	75 mg
			d.H_2O	Up to 500ml
			Vitamin Solution II	
			Thiamine hydrochloride (T4625)	25 mg
			Riboflavin (R4500)	25 mg
			d.H_2O	Up to 500 ml

Disk concentrations were selected based on advice for *Enterococcus* from CLSI, EUCAST and BSAC as a Gram-positive, facultative anaerobic Firmicute (Table 2.1). Though Bacteroidetes are Gram-negative, the same concentration disks were used for consistency. Zone of inhibition diameters were measured using a digital caliper after 48 hours and averaged across the biological replicates. The identity of each culture was confirmed using full-length PCR of the 16S rRNA gene (7F forward primer (5'-AGAGTTTGATYMTGGCTCAG-3') and 1510R reverse primer (5'-ACGGYTACCTTGTTACGACTT-3')) with the following program: 95 °C 15 mins; 35 cycles of 95 °C 30 s, 58 °C 30 s, 72 °C 2 mins; 72 °C 8 mins. DNA purification and capillary sequencing was performed by Eurofins Genomics (Germany). The forward and reverse sequences were trimmed and those shorter than 400bp discarded. The remaining forward and reverse reads were merged using Merger (Emboss: 6.3.1[197]) and BLASTn was used to identify a closest taxonomic match against the 16S rRNA gene sequences of the HBC[151].

Table 2.2. Single-concentration antibiotic disks used for phenotypic susceptibility testing. Antibiotics were chosen based on clinical relevance and presence of genetic determinants of antibiotic resistance in the HBC genomes.

Antibiotic (part number)	Antibiotic class	Oxoid Disk Concentration (µg)	Based on guidelines
Amoxicillin (11952962)	Beta-lactams: penicillin	10	*Enterococcus*; CLSI 2015
Ceftriaxone (11963812)	Beta-lactams: cephalosporin	10	*Enterococcus*; EUCAST 2018
Ciprofloxacin (11499838)	Fluoroquinolones	10	*Enterococcus*; CLSI 2015
Erythromycin (10280243)	Macrolides	15	*Enterococcus*; CLSI 2015
Gentamicin (10299772)	Aminoglycosides	30	*Enterococcus*; EUCAST 2018
Metronidazole (11913972)	Nitroimidazoles	5	Miscellaneous; BSAC 2015
Tetracycline (11963872)	Tetracyclines	30	*Enterococcus*; CLSI 2015
Trimethoprim (10597083)	Dihydrofolate reductase inhibitor	5	*Enterococcus*; EUCAST 2018
Vancomycin (11974012)	Glycopeptide	30	*Enterococcus*; CLSI 2015

Density curves of average zone of inhibition diameters for each phylum and antibiotic was plotted using R[198] and ggplot2[199]. In addition, the interquartile range of average zone sizes in isolates with and without genetic determinants of resistance were determined using the Tukey method[190].

2.5 Defining a scale for categorising resistant/susceptible phenotypes

Guidelines for defining resistance and susceptibility exist for a range of pathogens, though are less well-characterised in anaerobes and Gram-positive bacteria. Moreover, they are used for categorising closely related taxa, usually at the species level. In this thesis I wanted to compare at higher taxonomic levels, therefore I used the phenotypic data that I generated to create a scale, considering isolates 'resistant' if the zone size was in the lower quartile or smallest 25 % of all zone sizes for a particular antibiotic. Similarly, isolates were considered 'susceptible' if the zone size was in the upper quartile or largest 25 % of all zone sizes for a particular antibiotic. Isolates with zone sizes in the middle 50 % require further testing to determine whether or not they should be considered susceptible or resistant (see Fig. 4.4, Table 4.1 and Table 4.3 in Chapter 4). In combination with the genotypic data, there are four key genotype/phenotype combinations of confirmed or unpredicted susceptibility or resistance. These genotype/phenotype combinations were then converted to a four-number system and visualised as a heatmap in iTOL[191] against the phylogeny of the 73 isolates (generated as described in section 2.1). In addition, the proportions of these categories were determined for both the antibiotics tested and in each phylum. The enrichment of each combination in antibiotics or phyla was determined using Fisher exact tests[192] and corrected using the Benjamini, Hochberg, and Yekutieli method[193] (q significant < 0.05). The proportion and

enrichment of Unpredicted Resistances per phylum was determined for each antibiotic in the same way. Proportions of genotype/phenotype combinations were compared between databases or prediction methods as described in section 2.2. to the proportions first calculated using ARIBA[160] with CARD[163,164] with two-proportion z tests plus Yates correction (q value significant < 0.05).

2.6 Further investigations of Unpredicted Resistance

All phenotyped HBC isolates were ranked from most ceftriaxone-resistant to least ceftriaxone-resistant (i.e., from smallest to largest mean ceftriaxone zone of inhibition). Each average ceftriaxone zone of inhibition size was then plotted next to each isolate in this ranked order. Isolates in the top five most ceftriaxone-resistant isolates without beta-lactam resistance genes or mutations and with unexpectedly large average ceftriaxone zones of diameter (determined as described in section 2.4 using the Tukey method[190]) were investigated further (*Bacteroides faecis* 18048_2#66 and *Lachnospiraceae nov.* 20287_6#18). The HBC core gene phylogeny visualised in iTOL[191] from section 2.1 was used to identify the most closely related isolates in the HBC to *Bacteroides faecis* 18048_2#66 and *Lachnospiraceae nov.* 20287_6#18. Average Nucleotide Identity (ANI) between the isolates of interest and close relatives was performed using FastANI[200] with default parameters. The Minimum Inhibitory Concentration (MIC) for ceftriaxone was determined using Biomerieux Etest strips. Isolates were purified, grown in YCFA broth culture and used to inoculate YCFA agar plates as described for disk testing in section 2.4. Etest strips (ceftriaxone gradient 0.016-256 µg/mL; Biomerieux part number 506618) were placed on the plates and MIC measured after 48 hours growth in anaerobic conditions as for disk tests described in section 2.4. The most closely related isolate

to *Lachnospiraceae nov.* 20287_6#18 that was susceptible to ceftriaxone was only 81 % similar by ANI and so these isolates were excluded from subsequent analyses.

Roary[201] was used to identify genes unique to the resistant *B. faecis* isolates (MIC > 256 µg/mL) and genes unique to the susceptible *B. faecis* isolates (MIC < 48 µg/mL) as well as genes shared by both sets. The NCBI Protein database was searched for "beta-lactamase" and the amino acid sequence of all results was downloaded. ShortBRED[194] was used to group these sequences into unique amino acid markers. ShortBRED was also used to search the genes unique to the resistant isolates for the amino acid markers of beta-lactamase related sequences. A range of similarity cut offs from 90 % to 25 % were tested. The highest cut off, 90 %, represents a high level of similarity that can be used to infer function. A single beta-lactamase was identified in the *B. faecis* isolate ("Group 2384") with 90 % identity. The presence of this gene in the 737 HBC isolates was predicted using ARIBA[160] as in section 2.2 and was found in 16 additional HBC isolates. Five of those isolates also contained other beta-lactam resistance genes or mutations and so were excluded from downstream analyses. The MIC of ceftriaxone for the remaining 11 additional Group 2384-positive isolates and their most closely related Group 2384-negative isolates in the HBC was determined using Etests as before. ANI analysis was also performed for these isolates as above. The nucleotide sequence for the Group 2384 gene was extracted from each phenotyped Group 2384-positive isolate, aligned using Muscle[202,203] (visualised in SeaView[204]) and a phylogenetic tree inferred using the General Time Reversible model with FastTree[187,188].

The Group 2384 candidate beta-lactamase was synthesised in a plasmid vector containing a chloramphenicol resistance gene using GeneArt (ThermoFisherScientific, plasmid pACYC184[205], construct ID 18ADVNOP). The construct was transformed into electrocompetent ElectroMAX DH10B T1 Phage-Resistant Competent *E. coli* Cells

(ThermoFisherScientific, part number 12033015, via electroporation according to the manufacturer's instructions and with assistance from Mr Matthew Dorman). The ceftriaxone MIC of the recipient *E. coli* strain was determined using Etests as earlier described. The transformed cells were grown on LB agar plates containing chloramphenicol (12.5 µg/mL) and ceftriaxone at a concentration of 256 µg/mL (representing the observed *B. faecis* phenotype) or 4 µg/mL (slightly above *the E. coli*'s initial MIC).

Bacteroides faecis 18048_2#66 and *Lachnospiraceae nov.* 20287_6#18 isolates with unpredicted ceftriaxone resistance were purified and Mr Mark Stares isolated genomic DNA using phenol:chloroform extractions. The genomic DNA was separated by size using a 1 % low-melting point agarose gel and fragments of approximately 40kb extracted using the Copy Control Fosmid Cloning Kit (Lucigen part number CCFOS059) according to manufacturer's instructions. These fragments were cloned into ceftriaxone susceptible *E. coli* using the Copy Control Fosmid Cloning Kit (Lucigen part number CCFOS059) according to manufacturer's instructions. The vector contains a chloramphenicol resistance gene; clones were selected for on LB agar containing chloramphenicol (12.5 µg/mL). The ceftriaxone MIC of the recipient *E. coli* strain was determined using Etests as earlier described. The transformed *E. coli* were screened for gain of ceftriaxone resistance by growth on LB agar containing ceftriaxone at a concentration of 256 µg/mL (representing the observed *B. faecis* phenotype) and 4 µg/mL (slightly above the *E. coli*'s initial MIC).

2.7 Humanised microbiota mouse experiments

A mouse line with microbiota derived from a healthy human had been established previously by Dr Simon Clare, Dr Sam Forster, Dr B. Anne Neville and colleagues in the Wellcome Sanger Institute Research Support Facility, by oral gavage of homogenised stool into germ-free (GF) mice ("Donor 2 humanised microbiota mice"). A second mouse line with microbiota derived from a different healthy human had also previously been established using the same techniques ("Donor 7 humanised microbiota mice") by the same persons. Full details are described in section 5.2.1.

Mice were given a theoretically therapeutic dose (approximately 45 mg/kg/day), based on the concentration required to adequately exceed the MIC of sensitive organisms in otitis media infections[206] and assuming that the average mouse weighs 30 g and consumes approximately 5 ml of water per day (according to John Hopkins University, http://web.jhu.edu/animalcare/procedures/mouse.html, accessed June 2015 and June 2019, and as advised by Dr Simon Clare). Amoxicillin sodium (TOKU-E part number A059) was dissolved in water, sterilized using a 0.2 µm filter and given to the mice via drinking water for seven days by Dr Simon Clare and his team. Faecal pellets were collected from each mouse at various time points before and after treatment by Dr Simon Clare and his team. At each time point, I weighed each individual faecal pellet and homogenised them in 100 mg/ml in sterile PBS; faecal homogenates were pooled per cage. An aliquot of the undiluted pooled homogenate was treated with ethanol for 30 minutes (1:3 volumes of 70 % v/v ethanol) to select for ethanol-resistant and spore-forming organisms[146], and washed by centrifugation for 13200 g for five minutes at room temperature, before removal of the supernatant and resuspension in four volumes PBS. The wash was repeated twice more and after the third wash and removal of supernatant, the sample was resuspended in the original volume of PBS.

Untreated and ethanol-treated pooled homogenates were serially diluted 1 in 10 from 10^{-1} to 10^{-7}. Untreated dilutions were plated on modified complex, broad-range YCFA[188] (Table 2.1), with or without amoxicillin added, under aerobic conditions (37 °C in a New Brunswick Scientific Innova 42 incubator) and anaerobic conditions (37 °C in a Whitley DG250 workstation. Amoxicillin was included in the agar at a concentration representative of clinical resistance according to EUCAST and CLSI guidelines for anaerobic bacteria (8 mg/L). Ethanol-treated dilutions were plated as above except that the agar plates also contained the bile salt sodium taurocholate (STC; Fisher Scientific UK Ltd part number 10629452) to promote germination of spores[146].

2.8 Colony count data

At each time point, the number of colonies growing in each condition from each cage was counted and converted to colony forming units (CFU) per gram of stool. CFU/g values were averaged across both experiments for each culture condition: aerobic plates without amoxicillin, anaerobic vegetative plates without amoxicillin, anaerobic spore-forming plates without amoxicillin, aerobic plates with amoxicillin, anaerobic vegetative plates with amoxicillin and anaerobic spore-forming plates with amoxicillin. Standard deviation was determined for each of these six conditions. The colony count data was then tested for normal or log-normal distribution using the Anderson-Darling, D'Agostino and Pearson, Shapiro-Wilk, and Kolmogorov-Smirnov tests. As many of the conditions did not pass the normality tests, a non-normal distribution was assumed for performing Kruskal-Wallis (One-way ANOVA of non-parametric data) tests to determine statistically significant changes in bacterial load over the time course of the mouse experiments (comparisons were performed between time points).

The two-stage step-up method of Benjamini, Krieger and Yekutieli[193] was used to correct for multiple comparisons.

2.9 Isolation of individual isolates and analysis

Non-confluent colonies were picked from culture plates, after 24-hour and 48-hours of growth in aerobic and anaerobic conditions respectively, directly into 300 µl YCFA broth. This was done with the assistance of Dr Hilary Browne, Dr Sam Forster, Dr Elisa Viciani and Dr Ana Zhu alongside other members of the Host-Microbiota Interactions Laboratory. The aerobic colonies were left to grow in broth for another 24 hours; anaerobic colonies were left for 48 hours. Half a millilitre of each broth culture was mixed with 0.5ml of 50 % glycerol to create glycerol stocks and stored at -80 °C. The full-length 16S rRNA gene was amplified from each isolate using PCR and sequenced by Eurofins as described in section 2.4. Reads were aligned using ssu-align v0.1.1[207] and the STK alignment converted to MSA. Consensus OTU sequences were aligned with Mafft[185,186] and a phylogeny inferred using FastTree[187,188] for visualisation with ITOL[191]. The tree leaves were annotated based on the BLASTn results and Operational Taxonomic Units (OTUs) were assigned with Mothur v1.35.1[208] with a defined cutoff of 0.01 (i.e., 99 % similarity). The number of isolates in each OTU was visualised using Krona[189]. Isolates of interest were purified and genomic DNA extracted using phenol:chloroform by Mr Mark Stares. Sequencing was performed by the Sanger Sequencing Pipelines (WSI) team using Illumina X10 technology, generating 150 bp paired-end reads.

Paired-end sequence reads were filtered and assembled using the pipeline as earlier mentioned[183]. To perform phylogenetic analyses, the amino acid sequences of 40 core genes were extracted, aligned and visualised as for the HBC[184-188,191] in section 2.1. Each genome was

searched for the presence of known antibiotic resistance determinants described in CARD[163,164] using ARIBA[160] with default parameters, as for the HBC in section 2.2. The ARIBA results were compared between genomes from isolates cultured on agar with amoxicillin after the mice received amoxicillin and genomes of the same strain (>99 % 16S identity) from agar without amoxicillin before treatment. The sequence of the *cfxA* beta-lactamase was identified in isolates of one OTU, *Odoribacter splanchnicus* 99.556 %, after amoxicillin therapy but was not present in isolates of the same OTU from before therapy. The nucleotide sequences of *cfxA* were extracted from all the whole genomes generated in this study (where present) and aligned using Muscle[202,203]. The amoxicillin MIC was determined for four isolates of *Odoribacter splanchnicus* 99.556 %, as described in section 2.6 (amoxicillin gradient 0.016-256 µg/mL; Biomerieux part number 500918).

For each candidate OTU, Roary[201] was used to identify genes found in the isolates from culture plates containing amoxicillin after mice were treated with amoxicillin but absent from isolates in the same candidate OTU from culture plates without amoxicillin before the mice received treatment. Genes meeting this criteria were searched for sequences with 90 % similarity to the ShortBRED[209] beta-lactamase markers developed in section 2.6. In addition, paired-end sequencing reads from bacteria isolated after amoxicillin treatment were mapped against assembled contigs from an isolate of the same OTU cultured before amoxicillin treatment using Smalt (https://www.sanger.ac.uk/science/tools/smalt-0). Variant bcf files were produced and used to identify SNPs in coding sequences (open reading frames annotated as such if over 100 bases) in Artemis[210]. Coding sequences containing SNPs were compared to known protein sequences using BLASTx.

2.10 Metascrape and metagenomic analysis

The total growth on a culture plate was mixed with 1 ml sterile PBS and scraped off ("metascrape") for total DNA extraction using the FastDNA Spin Kit for Soil (MP Biomedicals part number 1149200) according to manufacturer's instructions. Metagenomic sequencing was performed by the Sanger Sequencing Pipelines (WSI) team using Illumina HiSeq 4000 technology, generating paired-end reads of 150 bp. Metagenomic samples were multiplexed at 32 samples per lane. Metagenomic data was analysed using Kraken[211] with default parameters to compare the sequence reads to databases of reference genomes and assign taxonomy. Raw read counts per species were used to determine alpha and beta diversity using R scripts developed by Dr Kevin Vervier at each experimental time point for three culture conditions (aerobic, anaerobic vegetative and anaerobic spore-forming), on agar plates with or without 8 mg/L amoxicillin added. Statistical significance was determined by Mann-Whitney U tests[212], adjusted for multiple-testing with the Benjamini, Hochberg, and Yekutieli method[193]; q value significant < 0.05. Read counts were normalised per sample and the normalised values were averaged across samples for each of the three conditions (aerobic, anaerobic vegetative and anaerobic spore-forming) from plates containing amoxicillin. The species were then ranked by their mean relative abundance in each culture condition and the mean relative abundance of the top 10 most abundant species plotted as stacked bar charts, with the relative abundances of the remaining species grouped as "Other". Relative abundance of species was compared between consecutive time points and the start and end of the experiment using two-proportion z tests with Yates correction (q value significant when $q < 0.05$). Finally, the whole genome sequences generated in this study were searched for in the metascrape data to confirm the presence or absence of candidate OTUs of increased amoxicillin resistance on plates containing amoxicillin using Mash[213] with default parameters.

An identity cut off of 0.99 (99 %) was used to conclude that a particular strain was present in a metascrape sample.