Chapter 5. Modelling the development of antibiotic resistance *in vivo*

5.1 Introduction

In the previous two chapters antibiotic resistance genotypes and phenotypes in commensal human gut bacteria were investigated. In addition to determining the taxonomic distribution of genetic determinants of resistance and comparing antibiotic resistance genotype with phenotype, it is also important to understand the dynamics of how gut bacteria develop and spread antibiotic resistance in the human microbiota. Several *in vitro* studies have investigated the horizontal transfer of antibiotic resistance genes between closely related bacterial species, such as plasmids from *Salmonella* into *E. coli*[257] or from *Lactococcus* spp. into *Bacteroides*, *Bifidobacterium* and *Enterococcus* spp.[258], but results from *in vitro* experiments cannot easily be extrapolated to *in vivo* situations. *In silico* analysis of publicly available bacterial genomes has identified antibiotic resistance genes with high sequence similarity between human and animal gut bacteria and pathogens[152,259], indicating horizontal transfer, but this is indirect evidence. For direct evidence, we need to study a baseline level of antibiotic resistance in the human gut microbiome and observe how it can change in response to selective pressure caused by antibiotic therapy.

Studies with this goal have been performed before, such as those that use 16S rRNA gene sequencing to assess impact of antibiotics on the overall community[178,260-262]. For example, a study of combined amoxicillin and metronidazole on wild-type mice found that approximately 70 % of 16S rRNA gene sequences in antibiotic treated mice were Proteobacteria, compared to just 1 % in control non-treated mice[179]. In addition, prolonged reductions in overall diversity were observed even after antibiotic treatment was stopped[179]. However, 16S rRNA gene

profiling studies do not allow for high-resolution taxonomic analysis, as often species share similar 16S rRNA gene sequences. Because of this, strains and some species (such as *E. coli* and *Shigella*) cannot be distinguished with this method[263]. Moreover, these studies rely on databases of characterised 16S rRNA gene sequences. As discussed previously, there are still uncharacterised human gut bacteria without reference genomes available, meaning not all 16S rRNA gene sequences can be classified taxonomically and are therefore difficult to study. This is also true of whole genome shotgun metagenomics, which relies on databases of published reference bacterial genomes: previous studies using whole genome shotgun metagenomics have identified high levels of unclassified sequences[264]. Reference genome based metagenomics, where a custom database of reference genomes is curated and tailored for metagenomic analysis relating to the study being performed, is an emerging method to circumnavigate these problems[151,265,266]. This involves culturing bacteria present in the sample to be studied, identifying novel bacteria and generating new reference genomes for these organisms, and combining these into a database of reference genomes[146,151]. Reference genome based metagenomics was implemented in the present study to provide this tailored, high-resolution analysis of antibiotic resistance dynamics in gut microbiota. The culturing and WGS component of reference genome based metagenomics also assists strain-level resolution, which has typically been difficult with whole genome shotgun metagenomics due to limitations with sequencing depth and reference databases[267].

Studies on humans are often confounded by factors such as different genetics, diets, lifestyles, health condition, infection state and being limited to relatively few participants. Experiments in mice therefore offer a much higher level of control, as we can use mice with the same genetic background, from the same litter, that are fed the same diet and kept in controlled living conditions, and it is possible to study large numbers of individuals in these controlled

conditions. Previous models have used the di-associated mouse, in which a recipient and donor strain of interest are introduced into GF mice to assess plasmid conjugation frequency in the gut environment. This model has indicated that inter-species transfer is possible through the transfer of tetracycline and erythromycin resistance from *Lactobacillus plantarum* to *Enterococcus faecalis*[268]. Conventional mouse models have helped identify that transfer of plasmids bearing antibiotic resistance genes (ARGs) can occur at high frequency in the gut[269], even when colonisation resistance is present. However, the di-associated model is considered the "worst case scenario" model due to the lack of colonisation resistance from indigenous microbiota and so does not model the typical gut environment[270]. Moreover, studies in wild-type mice are not fully representative of biological processes in humans as mice have gut microbiomes that differ from humans[271]. Our lab has established methods to colonize mice with human-derived gut microbiota: these are GF mice colonised by gut bacteria of healthy humans following faecal transplant of human stool into the GF mice. This provides a very powerful tool for controlled experiments regarding the gut microbiome in a manner that is more representative of processes in humans. In addition, by culturing and sequencing whole genomes of bacteria isolated before and after antibiotic exposure, there is tremendous power to discriminate ARG-HGT events at a level not previously possible.

In this chapter, I use mice with humanised gut microbiota to model dynamics of antibiotic resistance in the human gut bacterial community and individual strain evolution following antibiotic therapy. Amoxicillin was used as the antibiotic providing selective pressure on the gut bacteria of these humanised microbiota mice. Amoxicillin is considered a relatively broad-spectrum penicillin, and is one of the most prescribed drugs globally[18]. Moreover, amoxicillin is often administered orally, and so can directly impact on our indigenous gut microbiota as well as the infections it is prescribed to treat. As we are still discovering novel and

uncharacterised members of the human gut microbiota, the full impact of broad-spectrum antibiotics such as amoxicillin on commensal gut bacteria remains to be defined. To investigate the comprehensive impact of amoxicillin on human gut bacteria including those considered uncharacterised, I combined large-scale culturing, reference genome based shotgun metagenomics and whole genome sequencing with experiments in humanised microbiota mice. Large-scale culturing of thousands of strains of human gut bacteria allows the type and extent of any genomic and phenotypic changes caused by amoxicillin therapy to be investigated by actively tracking the movement of genes conferring AMR. Therefore, this chapter aims to assess the impact of amoxicillin therapy on amoxicillin resistance in gut bacteria of mice colonised with human-derived intestinal microbiota, uniquely at both the community level and genome level of individual strains.

5.2 Results

5.2.1 Overview of mouse models

To model the impact of amoxicillin on antibiotic resistance in the human gut microbiome, I utilised two mouse lines with human-derived gut microbiota that had been established in our laboratory prior to the start of my PhD. The mouse lines were established by performing faecal microbiota transplant (FMT) of homogenised stool from "healthy" human donors ("Donor 2" and "Donor 7", who also contributed faecal samples for developing the HBC[143]) into germ free mice. Donors were considered healthy if they had not taken antibiotics in the six months prior to donating stool, suffered any gastrointestinal conditions or taken oral medications. FMTs were performed via oral gavage to each GF mouse weekly for three weeks to establish Donor 2 and Donor 7 specific cohorts. This was repeated in several GF mice, which were then bred (e.g. Donor 2 mouse with Donor 2 mouse or Donor 7 mouse with Donor 7 mouse) to produce

two separate mouse lines. Culturing has been performed extensively on samples from human Donor 2 and Donor 7[146,151]. The gut microbiota of Donor 2 mice differs from Donor 7 mice in the following ways: different community composition (i.e., different bacteria, see Appendix 3, Fig. A3.1); different antibiotic resistance potential, including amoxicillin-resistant organisms and antibiotic-resistance genes present in each community (Appendix 3, Figure A3.2 & A3.3); colonisation resistance.

Mice from each line were given amoxicillin orally via drinking water for seven days. Faecal samples were collected immediately before and after amoxicillin treatment. The amoxicillin was provided at a theoretically therapeutic dose (approximately 45 mg/kg/day), based on the concentration required to adequately exceed the minimum inhibitory concentration (MIC) of sensitive organisms in otitis media ear infections[206]. The quantity and concentration of drinking water containing amoxicillin was determined using the average weight of a mouse (30 g) and approximate volume of water consumed per mouse per day (5 ml). I performed deep bacterial culturing with three conditions (aerobic, anaerobic, and targeted for spore-formers) on YCFA agar plates. YCFA is considered a medium able to culture a broad range of bacteria. Metascrape samples (metagenomic sequencing of the total bacterial growth on a culture plate) from vegetative bacterial growth on YCFA plates in anaerobic conditions have been shown to be representative of the community in the original stool sample that was cultured from[146]. Thus, we can use anaerobic vegetative metascrape samples to represent the overall gut bacteria community, as well as targeted metascrapes for particular phenotypes such as antibiotic resistance. Each condition was plated in duplicate: with and without the addition of amoxicillin at a concentration of 8 mg/L in the agar plates. This concentration represents a MIC level where if some pathogenic bacteria are still able to grow in its presence they are considered clinically resistant according to EUCAST and CLSI guidelines; thus, this

concentration is used to select what can be considered amoxicillin-resistant bacteria. Details of each experiment performed are described and summarised below (Fig. 5.1, Table 5.1).

In the initial experiment ("D7AMX1"), eight Donor 7 humanised microbiota mice were divided into two cages. Both cages received a therapeutic dose of amoxicillin as described above. Two faecal pellets were collected from each mouse on experimental days 0 (immediately before amoxicillin regime started), 3, 7 (amoxicillin regime ended), 10, 14 (seven day recovery) and 35 (28 day recovery). Both pellets per mouse were weighed, then one was immediately frozen at -80°C and the remaining pellets were pooled per cage. Pooled pellets were homogenised 100 mg/ml in sterile PBS and serially diluted 1 in 10, from 10^{-1} to 10^{-7}. Dilutions were plated as described above (Fig. 5.1). In this experiment, individual colonies were isolated (Table 5.1).

The Donor 7 humanised microbiota mouse experiment was repeated ("LJP01"), but instead of isolating individual colonies metascrapes were collected (Table 5.1). In this experiment, three cages each containing six Donor 7 humanised microbiota mice were treated with amoxicillin as before. Faecal samples were collected and processed as described for D7AMX1. Colonies were counted for all samples, conditions and time points. Bacterial growth on culture plates was mixed with 0.5 ml sterile PBS, scraped off each plate for Day 0, 7, 14 and 35 and collected for whole genome shotgun metagenomic extraction and sequencing.

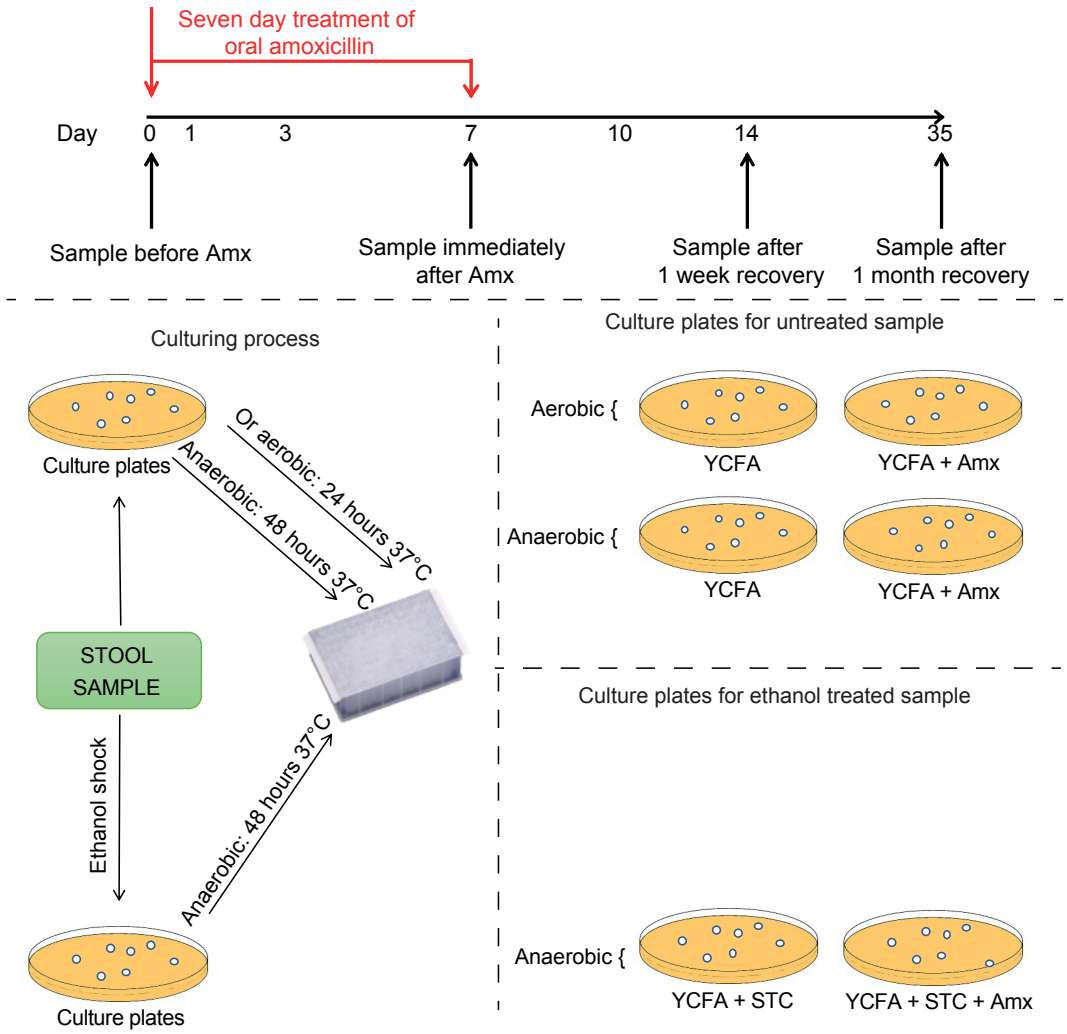

Figure 5.1. Diagram of experiments assessing the impact of amoxicillin on mice with human-derived gut microbiota. The timeline illustrates the sampling days of the experiments and the duration of amoxicillin treatment. The experiments were performed in both Donor 2 and Donor 7 humanised microbiota mice. The lower diagram illustrates the sample processing and different culture conditions and culture plate set ups: faecal pellets were collected from each mouse at each experimental time point. Faecal pellets were weighed and homogenised 100 mg/ml in sterile PBS, then pooled in equal amounts per cage. Pooled faecal homogenates were diluted 1:10 from 10^{-1} to 10^{-7} and appropriate dilutions plated on YCFA agar plates, without and with amoxicillin added (final concentration 8 mg/L amoxicillin). Plates were incubated for 24 hours at 37°C in aerobic conditions, or 48 hours at 37°C in anaerobic conditions. In addition, aliquots of the pooled faecal homogenates were ethanol shocked (diluted 1:4 in 70 % ethanol for 30 minutes) to kill vegetative bacteria and select for ethanol-resistant spores. Ethanol shocked samples were plated on YCFA plus sodium taurocholate (STC, a bile salt to promote spore germination) and incubated at 37°C in anaerobic conditions. All conditions were plated in duplicate to allow metascrape collection (collection of the total bacterial growth on agar plate for total DNA extraction and metagenomic sequencing) and the isolation of individual colonies. Individual colonies were picked into 96 well plates (one isolate per well) containing YCFA broth and incubated for 24 hours (aerobic) and 48 hours (anaerobic) at 37°C. The full length 16S rRNA gene sequence was amplified by PCR from each isolated colony in broth and sequenced by Sanger sequencing to assign species- or genus-level taxonomy based on similarity to reference 16S rRNA gene sequences.

A similar mouse model experiment was performed using three cages each containing six Donor 2 humanised microbiota mice (LJP02). The experiment set up was identical to that of LJP01 (Fig. 5.1) except that Day 0 samples were collected from all cages. In the LJP02 experiment all culturing plates were duplicated to allow both metascrapes to be collected and individual colonies to be isolated (Table 5.1).

Mouse line	No. mice	Sex	Colony counts	Isolated colonies for WGS	Metagenomics
Donor 2	18 across 3 cages	F	LJP02	LJP02	LJP02
Donor 7	18 across 3 cages	F	LJP01	×	LJP01
Donor 7	8 across 2 cages	M(4), F(4)	×	D7AMX1	×

Table 5.1. Summary of experiments assessing the impact of amoxicillin on mice with human-derived gut microbiota and samples or data generated. The table details the mice groups included in each experiment and rationale. Sex: F = female, M = male (brackets indicate number of each). The last three columns indicate what data was collected from which experiments (D7AMX, LJP01 or LJP02). × indicates not generated as part of that experiment.

5.2.2 Impact of amoxicillin on the bacterial load in mice with humanised gut microbiota

The total number of colonies growing on each plate for all pooled homogenates, conditions and time points were counted to determine changes in bacterial load over the time course of the LJP01 and LJP02 mouse experiments (Fig. 5.3). Raw colony counts were converted to colony forming units of bacteria per gram (CFU/g) of stool, representing bacterial load. The data was combined across the Donor 2 and Donor 7 experiments to look for general trends in bacterial load. The CFU/g data was then tested for normal or log-normal distribution using the Anderson-Darling[272], D'Agostino and Pearson[273], Shapiro-Wilk[274], and Kolmogorov- Smirnov tests[275]. As many of the colony count datasets did not pass the normality tests, a

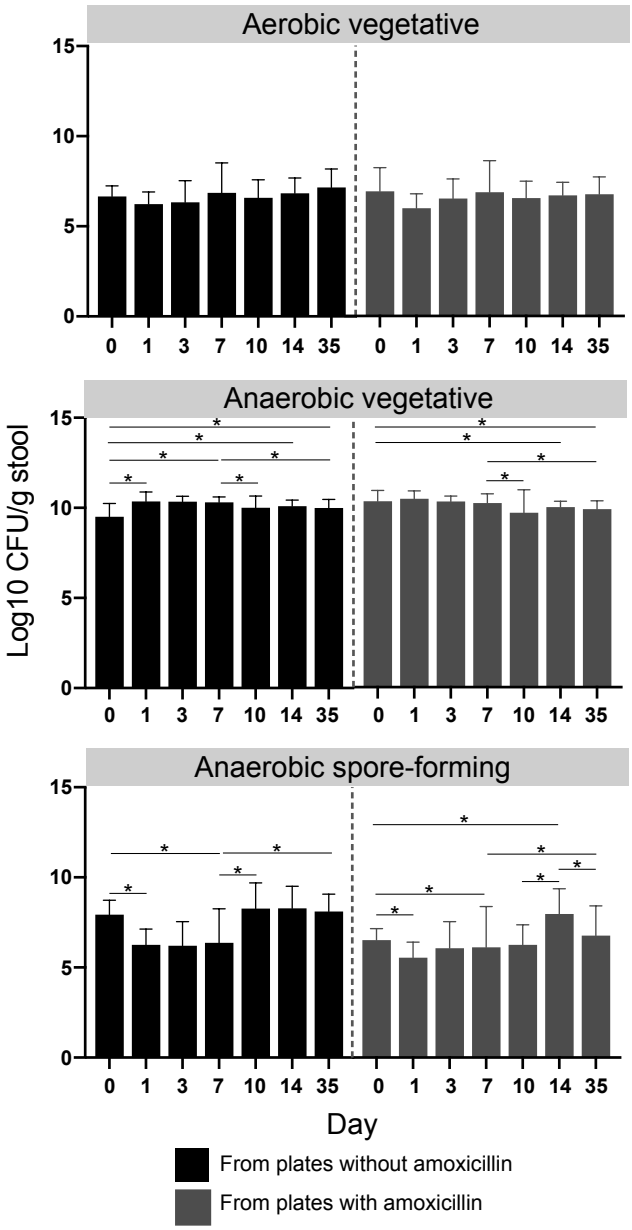

Figure 5.2. Bacterial load over the course of humanised microbiota mouse amoxicillin model experiments. Donor 2 and Donor 7 humanised microbiota mice were given a therapeutic dose of amoxicillin orally via drinking water for seven days; faecal pellets were collected per mouse, homogenized and pooled per cage per time point. The pooled homogenates were diluted and spread on agar plates in a variety of conditions: YCFA agar and anaerobic; YCFA agar plus 8 mg/L amoxicillin and anaerobic; YCFA agar, and aerobic; YCFA agar plus 8 mg/L amoxicillin and aerobic. Aliquots of each pooled homogenate was treated with 70 % ethanol for 30 minutes to select for spores, which were plated on YCFA agar + 1 % sodium taurocholate (STC) with and without 8 mg/L amoxicillin and grown anaerobically. Colonies were counted after 24 hours (aerobic growth) or 48 hours (anaerobic growth) and used to calculate the average bacterial load (colony forming units (CFU)/g stool) and standard deviation per cohort. A non-normal distribution was assumed for performing One-Way Anova (Kruskal-Wallis) tests to determine statistically significant changes in bacterial load over the time course of the mouse experiments (comparisons were performed between consecutive time points for all time points e.g. Day 0 vs. Day 1, Day 1 vs. Day 3, etc. and between major time points i.e., Day 0 vs. Day 7, Day 7 vs. Day 14, Day 14 vs. Day 35, Day 0 vs. 14, Day 0 vs. 35 and Day 7 vs. 35. The two-stage step-up method of Benjamini, Krieger and Yekutieli[193] was used to correct for multiple comparisons by controlling the false discovery rate. * marks where statistical significance ($q < 0.05$) was observed between marked time points.

non-normal distribution was assumed for performing Kruskal-Wallis (One-way ANOVA of non-parametric data) tests. Comparisons were performed between consecutive time points for all time points e.g. Day 0 vs. Day 1, Day 1 vs. Day 3, etc. and between major time points e.g. Day 0 vs. Day 7, Day 7 vs. Day 14, Day 14 vs. Day 35, Day 0 vs. Day 14, Day 0 vs. Day 35, Day 7 vs. Day 35. The two-stage step-up method of Benjamini, Krieger and Yekutieli[193] was used to correct for multiple comparisons by controlling the false discovery rate (q value significant < 0.05). There were no significant changes to aerobic bacterial load.

There was a significant increase in anaerobic vegetative bacterial load from plates without amoxicillin after the first day of treatment (Day 0 vs. Day 1, q value < 0.0001) and over the course of the amoxicillin therapy (Day 0 vs. Day 7, q value <0.0001). In addition, despite a significant reduction between Day 7 and Day 10 (q value 0.0124), the anaerobic vegetative load at recovery time points remained significantly higher than Day 0 (vs. Day 14 and Day 35, q values 0.0087 and 0.0201 respectively). However, Day 35 was significantly lower than Day 7 (0.0087). This indicates that the total anaerobic vegetative load increased after amoxicillin treatment began and then decreased after the treatment was stopped, but not back to its original level. There was also a significant decrease in anaerobic vegetative bacterial load from plates containing amoxicillin between Day 7 and Day 10 (q value 0.0138), and overall between Day 7 and Day 35 (q value 0.0026). There was an overall significant decrease in anaerobic vegetative bacterial load from plates containing amoxicillin (Day 0 vs. Day 14 and Day 35, q values 0.0010 and 0.0002 respectively). This indicates an overall decline in the load of amoxicillin-resistant anaerobic vegetative bacteria during the experiment.

The load of anaerobic spore-forming bacteria cultured on both plates with and without amoxicillin significantly decreased between Day 0 and Day 1 (q values 0.0003 and < 0.0001 respectively). In addition, both sets of anaerobic spore-forming bacteria showed significant

decreases in load between Day 0 and Day 7 (q values 0.0019 and < 0.0001 respectively). There was a subsequent significant increase in anaerobic spore-forming bacteria from plates with without amoxicillin after treatment stopped (Day 7 vs. Day 10, q value < 0.0001). The load of anaerobic spore-forming bacteria from plates without amoxicillin were significantly higher at Day 35 than Day 7 (q value <0.0001), but not significantly different to Day 0. Anaerobic spore-formers from plates with amoxicillin significantly increased between Day 10 and Day 14 (q value 0.0003), and between Day 14 and Day 35 (q value 0.0076). The Day 14 level was significantly higher than Day 0 (q value 0.00206) but whilst Day 35 was significantly higher than Day 7 (0.0163), it was not significantly different to Day 0. This indicates that the amoxicillin had a significant impact on reducing the load of spore-forming bacteria during the antibiotic treatment but that this community recovered in terms of CFU/g after the treatment stopped.

5.2.3 Deep culturing and whole genome sequencing to improve taxonomic classification of metagenomic data

Having studied the gross impact on bacterial load, the impact of amoxicillin on the diversity of the gut bacterial communities in these mice was investigated using high resolution metagenomics that can resolve to species and subspecies level. I first classified the metagenomic sequence reads from the metascrapes (total bacterial growth collected from culture plates) using Kraken[211], comparing them to a database of publicly available reference genomes of gut bacteria (including the 737 from the HBC[151]; this database was created in December 2017 by Dr Sam Forster). To determine the taxonomic resolution, I calculated the percentage of reads in each sample that could be classified to species level using this Kraken

database and determined the average proportion of classified sequence reads for each culture condition (aerobic, anaerobic vegetative and anaerobic spore-forming; Figure 5.4). Although the aerobic and anaerobic vegetative metascrapes were relatively well classified (> 95 %), lower levels of classified sequence reads were observed for spore-forming metascrapes (71.9 %). This indicates that this database is not optimal for classifying all of the metagenomic data generated in this study and I therefore sought to culture and whole genome sequence additional isolates to improve the taxonomic classification and downstream analysis.

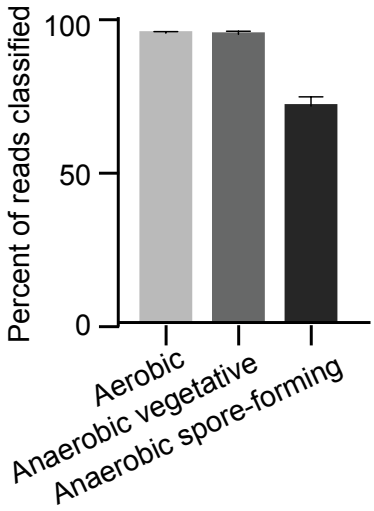

Figure 5.3. Proportion of classified metagenomic sequences reads from metascrapes of cultured faecal bacteria from mice with humanised gut microbiota. Sequence reads were assigned taxonomic classification using Kraken and a database containing publicly available gut bacterial and archaeal genomes, including those in the HBC[151]. The average percentage of classified reads for each culture condition was determined. Bars represent Standard Error of the Mean. This shows that this database was not capable of assigning taxa to all sequence reads, especially for spore-forming bacteria.

To identify candidates of novel reference genomes that might explain some of the unclassified sequence reads, I assembled contigs *de novo* from unclassified reads in one metascrape sample from the Donor 7 and one from the Donor 2 mouse experiments. I then performed a BLASTn analysis of those contigs against the NCBI RefSeq database (as of April 2017). Overall, the majority of unclassified contigs from the Donor 7 mouse sample were most similar to sequences belonging to Firmicutes, particularly the Clostridiales order. Over a quarter (27 %)

of the contigs had high similarity to a *Turicibacter* genome (Fig. 5.4A); this was also true of the Donor 2 mouse unclassified contigs (Fig. 5.4B). This indicated a *Turicibacter*-like organism could explain some of the unclassified data and would be a good candidate for a novel reference genome to add to the Kraken database.

Subsequently, I isolated approximately 11,000 colonies from across all the culture conditions to identify previously uncultured bacteria, including this candidate *Turicibacter*-like organism, and obtain as much diversity from the gut microbiota of these mice as possible. Capillary sequencing of the 16S rRNA gene was performed for all isolates. This resulted in 8838 full length 16S rRNA gene sequences being returned; the reduction in numbers of isolates to 16S rRNA gene sequences is likely caused by the failure to grow in broth or contamination of isolates. The 16S rRNA gene sequences were analysed and assigned OTUs with Mothur[208] using a 99 % identity cut off (i.e., sequences 99 % similar to each other were grouped into one OTU) in terms of similarity to the HBC[143] 16S rRNA gene sequences. In total, 367 OTUs were observed.

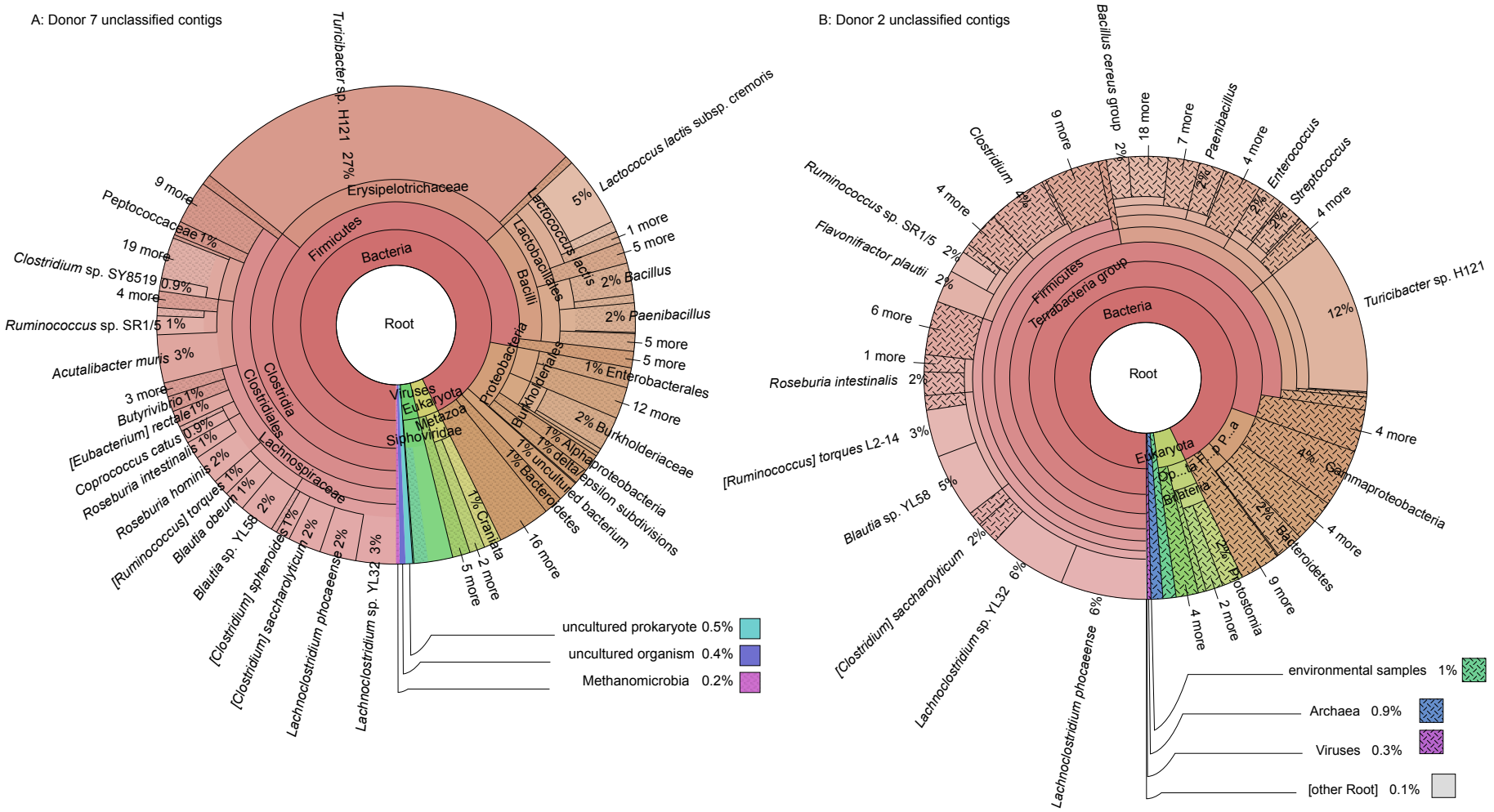

Figure 5.4. Diversity of the contigs assembled de novo from unclassified metagenomic sequence reads. *De novo* assembly was performed on the unclassified reads from a single metagenomic sample, with a high level of unclassified sequence reads, and the resulting contigs analysed using the NCBI RefSeq BLAST database containing all published genomes (as of April 2017). The chart shows the proportion of sequence reads by their most closely related reference genome (inner most ring: lowest taxonomic level e.g. domain; outer most ring: highest taxonomic level e.g. species). In this sample, the highest proportion of unclassified reads were most closely related to *Turicibacter sanguinis*, indicating that bacteria similar to this species would be useful to culture and generate new genomes from.

A rarefaction curve of this data was plotted: as the number of colonies picked increased, the number of OTUs isolated continued to increase even past 8000 colonies (Fig. 5.5). The trend was modelled to predict that the maximum OTU richness would feature 427 OTUs across the two mouse lines, but that approximately 30,000 colonies would be needed to be picked to recover all 427 OTUs. The 8838 colonies I isolated recovered approximately 86 % of the predicted total number of OTUs in gut microbiota of these mice.

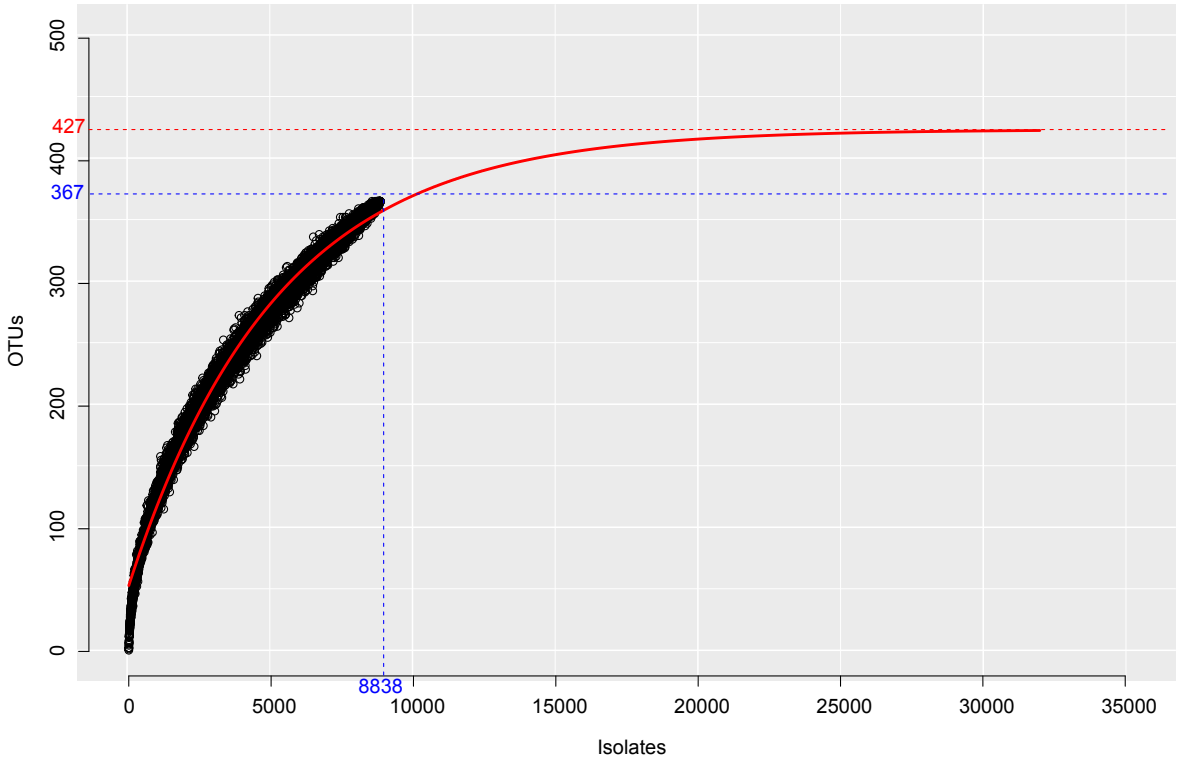

Figure 5.5. A rarefaction curve of the number of OTUs observed against the number of colonies picked. 11,000 individual bacterial colonies were isolated from mice with humanised gut microbiota before and after amoxicillin treatment. 8838 full length 16S rRNA gene sequences were generated and grouped into operational taxonomic units (OTUs) using a 99 % similarity cut off. The number of new OTUs observed as the number of colonies picked increases was plotted, showing that even as we reach almost 9000 colonies new OTUs are still being obtained. The blue dotted lines mark the maximum numbers of isolates picked and number of OTUs observed. The trend in part A was modelled using nonlinear regression to produce a rarefaction curve (solid red line) and estimate the maximum number of OTUs likely to be found across the two mouse lines with humanised gut microbiota: the red dotted line represents this predicted maximum of 427 OTUs. It would be necessary to pick approximately 30,000 colonies to recover the 427 OTUs. Dr Hilary Browne, Dr Sam Forster, Dr B. Anne Neville, Mr Mark Stares, Dr Elisa Viciani and Dr Ana Zhu all helped pick colonies from culture plates into broth cultures.

The consensus 16S rRNA gene sequence for each OTU was compared to 16S rRNA gene sequences in NCBI RefSeq (April 2017); the majority belong to Firmicutes (54.8%, Fig. 5.6) and Bacteroidetes (33.8%), with just a few Proteobacteria (11.4%) and Actinobacteria (1.36%).

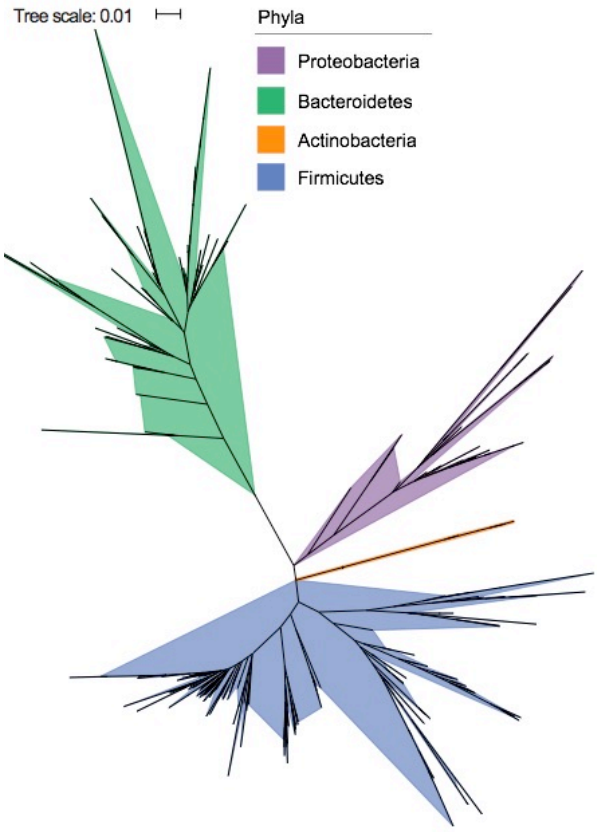

Figure 5.6. A phylogeny of the consensus sequences of 367 OTUs identified from 8838 full length 16S sequences generated in this study. 11,000 individual bacterial colonies were isolated from mice with humanised gut microbiota before and after amoxicillin treatment. 8838 full length 16S sequences were generated and grouped into operational taxonomic units (OTUs) using a 99 % similarity cut off. Consensus sequences of the 367 OTUs were compared to NCBI RefSeq (April 2017). The majority of the OTUs belong to the Bacteroidetes and Firmicutes phyla.

Looking at the proportion of colonies picked per OTU shows that approximately 1100, or 13 %, of the 8838 16S rRNA gene sequences were clustered in an OTU with 97 % nucleotide identity to *Turicibacter sanguinis* (Figure 5.7A). 16S rRNA gene sequences most similar to *Turicibacter* also demonstrated variability (Figure 5.7B), showing that these isolates are not all identical. Three isolates related to *Turicibacter* and 49 others with less than 97 % identity

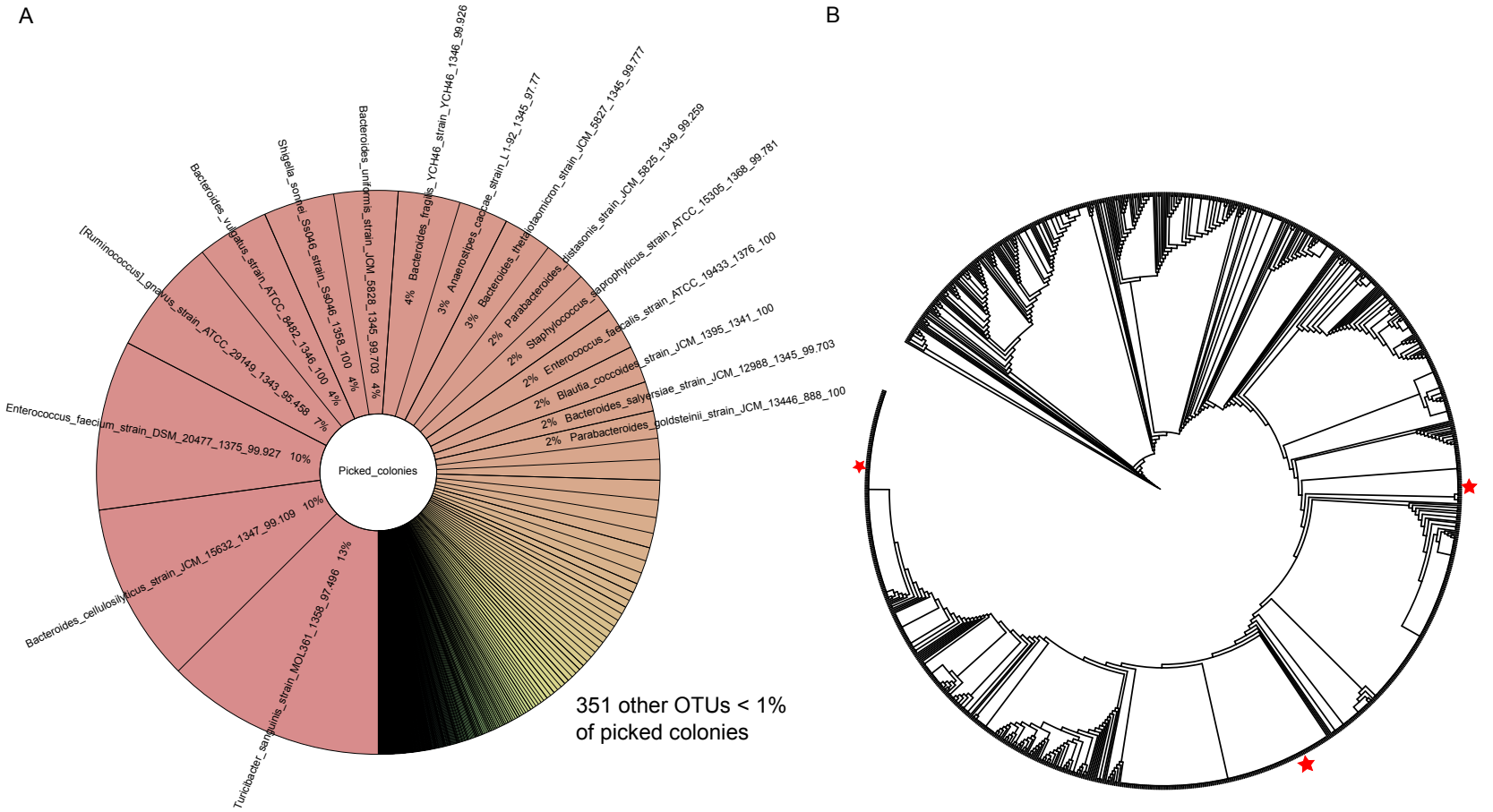

Figure 5.7. Diversity of OTUs isolated by culture from mice with human-derived microbiota. A) 16S rRNA gene amplification and Sanger sequencing was performed for all bacteria isolated in the mouse experiments described in this chapter, resulting in 8838 16S full length sequences. These were grouped into OTUs using Mothur and a 99 % cut off (i.e., 16S rRNA gene sequences >99 % similar to each other were considered one OTU); consensus OTU sequences were generated as part of this process. OTUs were named after the reference genome of highest similarity plus the length of the sequenced 16S rRNA gene and the percentage nucleotide identity to the reference sequence. For example, *Turicibacter_sanguinis*_strain_MOL361_1358_97.496 is the name of an OTU whose 16S rRNA gene is 1358 bases in length and 97.496% similar to the 16S rRNA gene sequence of *Turicibacter sanguinis* strain_ OL361. The proportion of all isolates each OTU contributed was calculated: 1127, or 13 %, were clustered in an OTU with 97 % identity to *Turicibacter sanguinis.* B) A phylogenetic tree of approximately 1127 16S rRNA gene sequences with 97 % identity to the reference *Turicibacter sanguinis* strain MOL361 genome shows there is considerable diversity even within the taxon. The red stars mark three isolates that were purified and DNA extracted for WGS.

to known 16S rRNA gene sequences (and therefore potentially novel species) were purified and whole genome sequenced. In addition, 11 isolates representing very common OTUs (>20 colonies isolated per OTU) and 135 isolates representing taxa that appeared to have increased in amoxicillin resistance during the experiment were purified and whole genome sequenced. Their relationship with the HBC isolates is demonstrated in Figure 5.8.

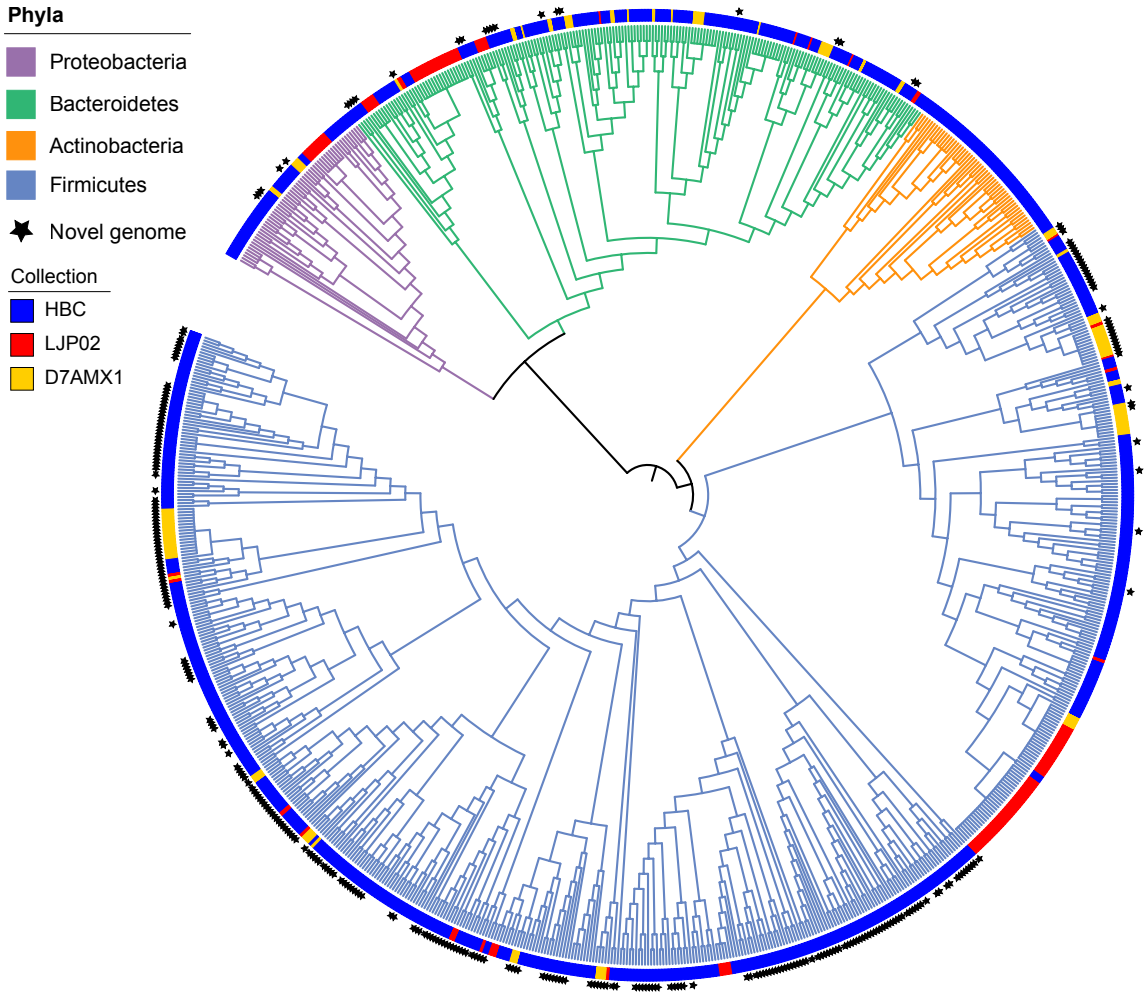

Figure 5.8. The phylogenetic relationship between the HBC and 198 new genomes generated in this study. The Human Gut Bacteria Culture Collection currently contains 737 isolates of human gut bacteria. The amino acid sequences of 40 core genes were extracted from all genomes and used to infer a phylogeny to illustrate the taxonomic diversity of the HBC and 198 genomes generated in this thesis. The stars mark which genomes are considered 'novel' based on the similarity of their 16S rRNA gene sequence to known 16S rRNA gene sequences in RefSeq. Mr Mark Stares performed genomic DNA extractions for the 198 D7AMX1 or LJP02 isolates.

All 198 new genomes were added to the Kraken database used earlier for reclassification of the metagenomic sequence reads from the metascrape data. The custom database contains publicly available gut bacteria genomes plus the 198 genome assemblies generated in this study (of which 52 represent isolates of potentially novel new taxa (16S ID < 98.7 %). The classification of metagenomic sequence reads to species level was improved from 95.6 % to 99.1 % in aerobic metascrapes and 95.3 % to 96.5 % in anaerobic vegetative metascrapes (Figure 5.9). The spore-forming samples were still less well classified at 86.6 % but also the most improved (up from 71.9 %). The resolution of downstream metagenomic data analysis was higher as a consequence.

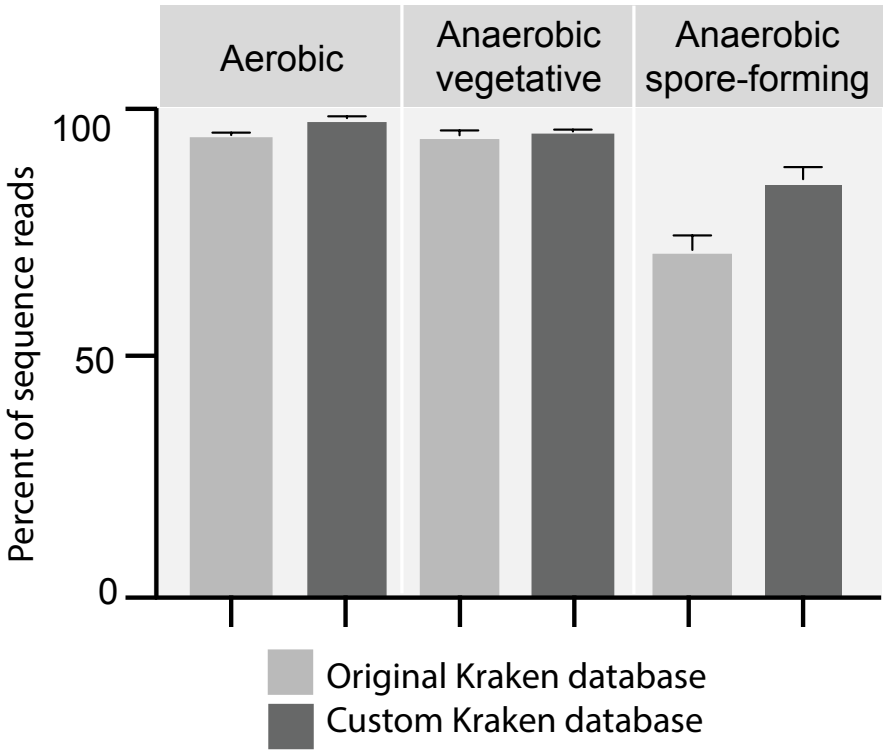

Figure 5.9. Comparisons of proportion of classified metagenomic sequences reads from metascrapes of cultured faecal bacteria from mice with humanised gut microbiota using different databases of reference bacterial genomes. Sequence reads were assigned taxonomic classification using Kraken and a database containing publicly available gut bacterial and archaeal genomes, including those published by Forster *et al.* 2019[151] (the "original" database), plus the same collection with new genomes generated in this thesis added ("custom"). The average percentage of classified reads for each condition was determined. Bars represent Standard Error of the Mean. This shows improved classification rates for all conditions when using the custom database compared to using the original database.

5.2.4 Impact of amoxicillin on the amoxicillin-resistant community

The updated custom Kraken database was used to assign taxonomy to sequence reads and assess the impact of amoxicillin on the diversity of the amoxicillin-resistant community in humanised microbiota mice. Firstly, the raw read counts assigned to each species in the Kraken database were used to determine the Shannon index of alpha diversity. Alpha diversity indices were averaged over metascrape samples from plates containing 8 mg/L amoxicillin, combined across both the Donor 2 and Donor 7 experiments, for each of the three culture conditions (i.e., aerobic, anaerobic vegetative and anaerobic spore-forming from plates containing amoxicillin).

Mean alpha diversity increased in aerobic and anaerobic vegetative metascrapes from plates with amoxicillin (Fig. 5.10) between Day 0 and Day 7, then decreased again after amoxicillin treatment had stopped, though remained higher than initial levels. This may indicate that the amoxicillin caused additional species to become resistant to amoxicillin within an individual, although this was not maintained after the treatment stopped. Alternatively, it may be that amoxicillin-resistant bacteria expanded in abundance from below detection level to above detection level, or that certain species of the same relative abundance were harmed by the amoxicillin, allowing the remaining species of that initial relative abundance to become detectable. In contrast, anaerobic spore-forming metascrapes from plates with amoxicillin, alpha diversity increased over all time points. This indicates that more spore-forming species became resistant over the experiment, or became more detectable, even after the treatment was stopped. However, there was no significant difference between any consecutive time points (i.e., Day 0 vs. Day 7 , Day 7 vs. 14, Day 14 vs. 35), nor between the start and end of the experiment (Day 0 vs. 35) for any culture condition (statistical significance determined by

Mann-Whitney U tests[212], adjusted for multiple-testing with the Benjamini, Hochberg, and Yekutieli method[193]; q value significant < 0.05).

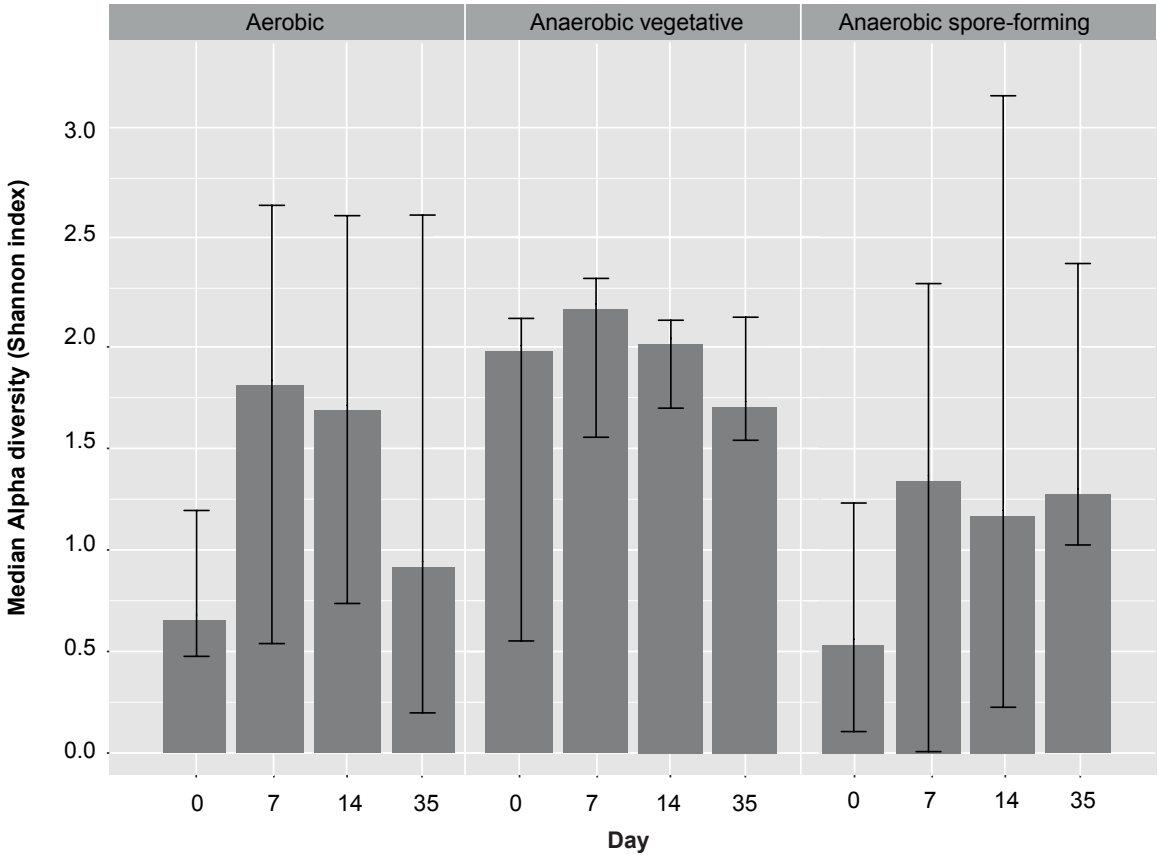

Figure 5.10. Changes in alpha diversity in metascrapes of cultured faecal bacteria from mice with humanised gut microbiota treated with amoxicillin. Sequence reads were assigned taxonomic classification using Kraken and a database containing publicly available gut bacterial and archaeal genomes plus genomes generated in this study. Raw read counts per species were used to determine alpha diversity at each experimental time point (Day 0, 7, 14, 35) for three culture conditions (aerobic, anaerobic vegetative and anaerobic spore-forming), on agar plates with 8 mg/L amoxicillin added. The median alpha diversity increased in aerobic metascrapes from plates with amoxicillin during the amoxicillin treatment (between Day 0 and Day 7) and reduced after amoxicillin treatment had stopped (Day 14 and Day 35), but remained higher than its initial level. A similar pattern was seen in anaerobic vegetative metascrapes from plates with amoxicillin but to a lesser extent. In anaerobic spore-forming metascrapes from plates with amoxicillin, where alpha diversity increased over all time points. There were no significant differences between consecutive time points (i.e., Day 0- vs. , Day 7 vs. 14, Day 14 vs. 35), nor between the start and end of the experiment (Day 0 vs.35) for any culture condition. Statistical significance determined by Mann-Whitney U tests[212], adjusted for multiple-testing with the Benjamini, Hochberg, and Yekutieli method[193]; q value significant < 0.05. Error bars represent the range. Dr Kevin Vervier wrote the R scripts that measured and compared alpha diversity.

Beta diversity (Fig. 5.11) was also determined to identify whether individual samples (i.e., metascrapes from a particular cage of mice) changed in similar ways during the experiment. Bray-Curtis dissimilarity measures were determined from raw read counts and averaged over metascrape samples from plates containing 8 mg/L amoxicillin, across both experiments, for each of the three culture conditions (i.e., aerobic, anaerobic vegetative and anaerobic spore-forming from plates containing amoxicillin). Although in all culture conditions beta diversity increased over the time course of the experiment, indicating that the amoxicillin-resistant communities belonging to different cages of mice became slightly more different from each other following amoxicillin therapy (i.e., there was a cage-specific effect), again, there were no significant differences between consecutive time points (i.e., Day 0 vs. Day 7, Day 7 vs. 14, Day 14 vs. 35), nor between the start and end of the experiment (Day 0 vs. 35) for any culture condition (determined as described for alpha diversity). This indicates that overall, the amoxicillin-resistant communities belonging to different cages did not become significantly different from each other during or after amoxicillin treatment.

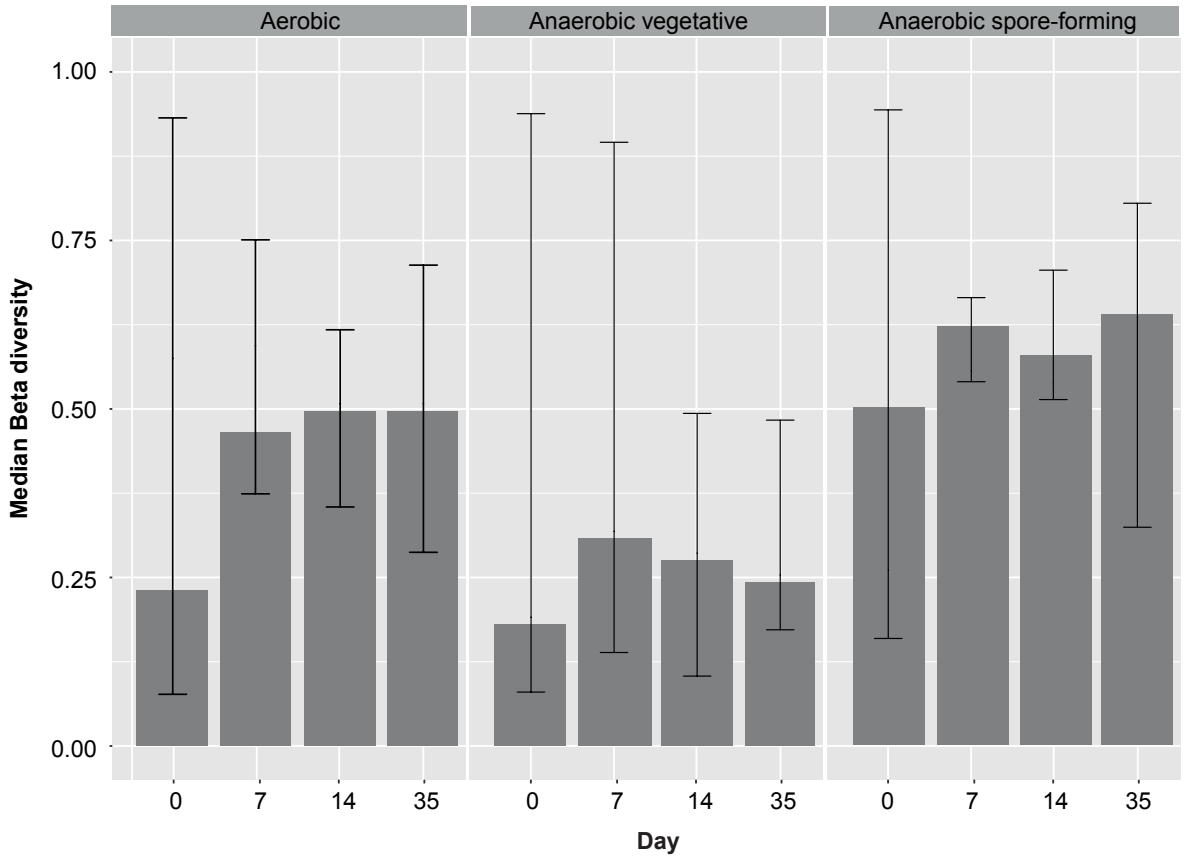

Figure 5.11. Changes in beta diversity in metascrapes of cultured faecal bacteria from mice with humanised gut microbiota treated with amoxicillin. Sequence reads were assigned taxonomic classification using Kraken and a database containing publicly available gut bacterial and archaeal genomes plus genomes generated in this study. Raw read counts per species were used to determine beta diversity at each experimental time point (Day 0, 7, 14, 35) for three culture conditions (aerobic, anaerobic vegetative and anaerobic spore-forming), on agar plates with 8 mg/L amoxicillin added. Median beta diversity increased in aerobic metascrapes over all time points. In both anaerobic vegetative and anaerobic spore forming metascrapes from plates without amoxicillin, beta diversity increased between Day 0 and Day 7, dropped slightly at Day 14 then increased again at Day 35. Anaerobic vegetative and anaerobic spore-forming metascrapes from plates with amoxicillin also show an increase in beta diversity from Day 0 to Day 7 which then decreases at both Day 14 and Day 35. There were no significant differences between consecutive time points (i.e., Day 0- vs. Day 7, Day 7 vs. 14, Day 14 vs. 35), nor between the start and end of the experiment (Day 0 vs.35). for any culture condition. Statistical significance determined by Mann-Whitney U test[212], p value significant < 0.05. Error bars represent the range. Dr Kevin Vervier wrote the R scripts that measured and compared beta diversity.

Despite the overall changes in diversity not being significant, it is possible that there were significant changes occurring in individual taxa. To assess this, the relative abundance of species in the amoxicillin-resistant communities for each culture condition (aerobic, anaerobic vegetative and anaerobic spore-forming metascrapes from plates containing amoxicillin) was determined. Read counts were normalised per sample and then the normalised values were averaged across all samples from the Donor 2 and Donor 7 experiments for each of the three

conditions. The species were then ranked by their mean relative abundance in each culture condition and the mean relative abundance of the top ten most abundant species plotted, with the relative abundance of the remaining species grouped as “Other”. This “Other” category formed the majority of (>92 %) the relative abundance of species in the anaerobic vegetative and spore-forming metascrapes from plates with amoxicillin (Fig. 5.12) at all time points.

For the aerobic amoxicillin-resistant community, *Enterococcus faecalis* dominated at Day 0 before amoxicillin therapy (mean relative abundance 78.5 % of sequence reads per sample) and was significantly reduced at Day 7 (0.88 % mean relative abundance, q value < 0.001; determined by unpaired t-tests with Welch’s correction[276], adjusted for multiple-testing with the Benjamini, Hochberg, and Yekutieli method[193] and significant < 0.05). The mean relative abundance of *E. faecalis* then increased by Day 14 (to 8.19 %), and was dominant again at Day 35 (80.0 %, significantly increased from Day 14; q value < 0.001). Overall, there was no significant difference between the initial (Day 0, 78.5 %) and final measured mean relative abundance of *E. faecalis* (Day 35, 80.0 %), showing that although this species reduced in relative abundance following amoxicillin therapy it recovered to its initial level after treatment stopped.

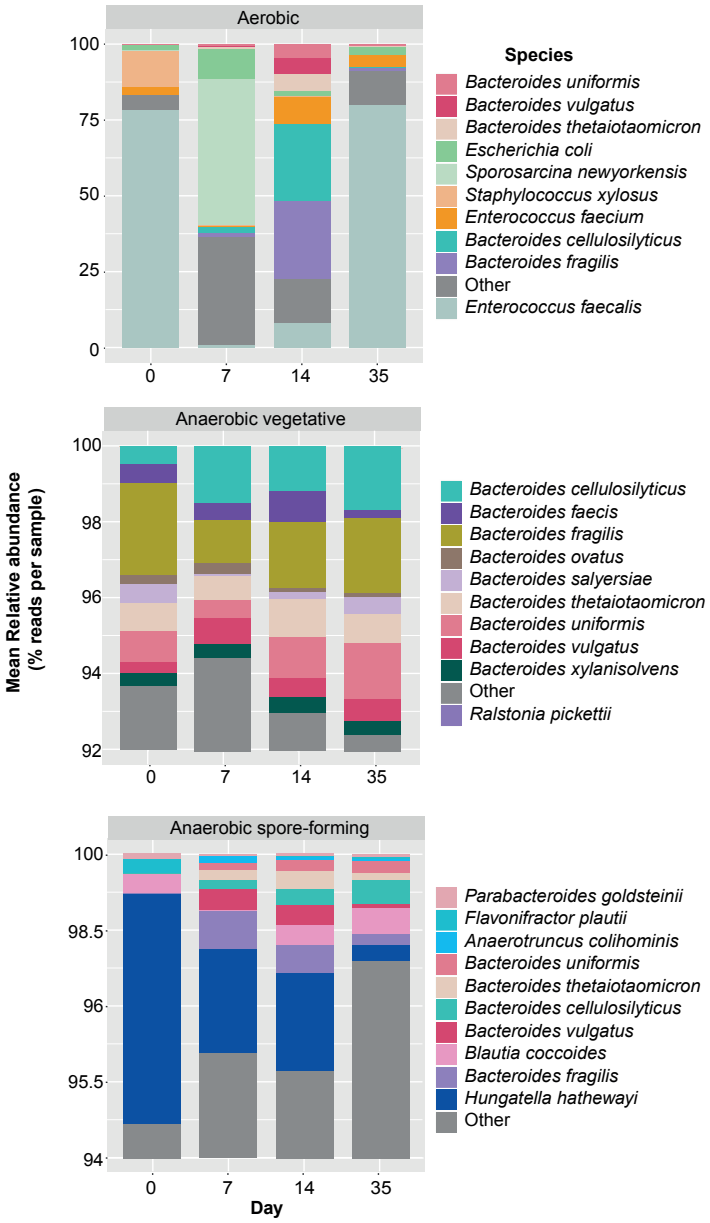

Figure 5.12. Relative abundance of species in metascrape samples cultured from mice following amoxicillin therapy. Sequence reads were assigned taxonomic classification using Kraken and a database containing publicly available gut bacterial and archaeal genomes plus genomes generated in this study. Read counts were normalized per sample and the relative abundance of each species averaged across all samples for each culture condition. Species were ranked from highest to lowest mean relative abundance per culture condition and the top 10 species for each are shown in these bar plots. All other species were grouped as "Other"; this category made up over 92 % of the total relative abundance of bacteria in the anaerobic vegetative and anaerobic spore-forming samples and so the y axis for these two conditions is limited to show the ten most abundant species more closely. Only metascrapes from plates containing amoxicillin are shown to represent the amoxicillin-resistant community. The aerobic amoxicillin resistant community was dominated by *Enterococcus faecalis* at the start of the experiment and end of the experiment (Day 0 and Day 35) but was much less relatively abundant at days 7 and 14, which were dominated by *Sporosarcina newyorkensis* or *Bacteroides fragilis* and *B. cellulosilyticus* respectively. The anaerobic vegetative amoxicillin resistant community most obviously shows a reduction in *Ralstonia picketti* and expansion of *B. cellulosilyticus* from Day 0 over all other time points, plus an increase in "Other" species at Day 7 that was not maintained at days 14 or 35. Similarly, the anaerobic spore-forming amoxicillin resistant community shows continual increases in "Other" species as well as a reduction in *Hungatella hathewayi*.

Whilst *E. faecalis* was less relatively abundant at Day 7 than Day 0, the amoxicillin-resistant aerobic community at Day 7 was dominated by *Sporosarcina newyorkensis: S. newyorkensis* was significantly higher at Day 7 (48.2 %) than Day 0 (0.17 %, q value < 0.001), but had reduced again by Day 14 (0.003 %, q value < 0.001). The relative abundance of *S. newyorkensis* was lowest at the end of the experiment (Day 35, 0.0005 %) but this was not significantly different from Day 14 or Day 0. Again, this shows that although this species increased in relative abundance following amoxicillin therapy it also returned to its initial level after treatment stopped. *Bacteroides fragilis* and *B. cellulosilyticus* were the dominant species at Day 14 (mean relative abundance 25.5 % and 25.3 % respectively). These two species were both significantly more relatively abundant at Day 14 than Day 7 (*B. fragilis* 1.2 %, *B. cellulosilyticus* 2.19 %) or Day 35 (*B. fragilis* 0.99 %, *B. cellulosilyticus* 0.47 %; q values all < 0.001). There was no significant difference between the initial and final relative abundance of *B. fragilis* and *B. cellulosilyticus* (Day 0: 0.16 % and 0.06 % respectively; Day 35: 0.99 % and 0.47 % respectively). This shows that these *Bacteroides* species became dominant in the aerobic amoxicillin-resistant community in the initial seven day recovery period following treatment, but had returned to their initial level by the end of the experiment. Finally, the relative abundance of "Other" species increased significantly from Day 0 (4.40 %) to Day 7 (29.6 %, q value < 0.0001) and significantly decreased again between Day 7 and Day 14 (4.97 %, q value < 0.0001). There was a slight increase in the relative abundance of "Other" species at Day 35 (9.96 %), but this was not significantly different from Day 14 or Day 0. This indicates that less abundant amoxicillin-resistant aerobic species became more abundant during the amoxicillin therapy and then returned to their initial level after the treatment ended. Overall, this suggests the amoxicillin therapy caused significant changes to the relative abundance of several amoxicillin-resistant species capable of growing in aerobic conditions, but that the final overall composition of this community was similar to its initial state.

The amoxicillin-resistant anaerobic vegetative community was less dominated by particular species: the mean relative abundance of "Other" species was consistently high (>92 % of sequence reads), though there was a slight increase at Day 7 (94.0 % compared to 92.2 % at Day 0) that was not maintained by Day 14 (92.6 %) or 35 (92.3 %). *Bacteroides fragilis* was the most relatively abundant species at 2.38 % of sequence reads per sample at Day 0; this reduced to 1.14 % at Day 7, then increased to 1.73 % at Day 14 and 2.00 % at Day 35. There was also a reduction in *Ralstonia picketti* (1.24 % at Day 0 to < 0.0001 % at Day 7, 14 and 35) and an increase of *B. cellulosilyticus* from Day 0 (0.45 %) over all other time points (1.47 % at Day 7, 1.16 % at Day 14, 1.67 % at Day 35). However, none of these changes were statistically significant. This shows that the relative abundance of individual species in the amoxicillin-resistant anaerobic vegetative community was not significantly affected by amoxicillin therapy.

In the amoxicillin-resistant anaerobic spore-forming community, *Hungatella hathewayi* was the most relatively abundant species at Day 0 (4.55 %) that steadily reduced at each time point (2.04 % at Day 7, 1.92 % at Day 14 and 0.33 % at Day 35). There was also a slight increase in "Other" species overall (94.6 % at Day 0, 96.0 % at Day 7, 95.5 % at Day 14 and 97.8 % at Day 35). However, none of these changes were statistically significant: the amoxicillin treatment did not have a major impact on the relative abundance of amoxicillin-resistant anaerobic spore-forming species. Overall, this suggests the amoxicillin therapy had more of an impact on the relative abundance of species in the amoxicillin-resistant aerobic community than the anaerobic vegetative or anaerobic spore-forming communities.

5.2.5 *Characterisation of strain- and sequence-level changes in gut microbiota following exposure to amoxicillin*

While metagenomics provides interesting insights of particular taxa within a community, individual isolates provide the opportunity to link genotype to experimentally validated phenotype. As earlier discussed, 8838 isolates with 16S rRNA gene sequences were obtained in these experiments and assigned to 367 OTUs. A record was kept of which culture plates and thus conditions, experimental time point and mouse line that each colony was picked from. This allowed the investigation of the impact of amoxicillin therapy on particular OTUs. The 16S rRNA gene sequence data was used to determine which OTUs were picked from agar plates containing 8 mg/L amoxicillin after the mice received antibiotic treatment (Day 7, 14 or 35) but not before (Day 0). OTUs that fit this criteria therefore represent lineages of bacteria that have either increased in relative abundance to above the detection limit or have increased in amoxicillin resistance (from under to over 8mg/L) during the course of the experiment. 14 OTUs fitted this criteria of only being picked from the amoxicillin agar plates after the mice received treatment. Isolates from before and after amoxicillin treatment in the mice of the 14 candidate OTUs of increased amoxicillin resistance were purified and whole genome sequenced (Table 5.2). These whole genome sequences were then searched for in the metascrape data from plates containing amoxicillin at Day 0 using Mash[213] to confirm the observations based on isolated colonies. All OTUs except *Bacteroides stercoris* 99.703 % and *Shigella sonnei* 99.757 % still fitted the criteria. For this reason, these two OTUs plus *Alistipes senegalensis* 98.49 %, *Alistipes shahii* 99.703 % and *Anaerostipes caccae* 99.773 % (isolates of which could not be recovered sufficiently for whole genome sequencing) are excluded from downstream analyses.

Table 5.2. Candidate OTUs of potential increase in amoxicillin resistance following *in vivo* exposure to amoxicillin. OTUs were considered candidates of increased amoxicillin resistance if they were only isolated from agar containing amoxicillin after the mice received amoxicillin (Day 7/14/35) but not before (Day 0). The number of isolates in each OTU at Day 0 (plates without amoxicillin), Day 0 (plates with amoxicillin) and the later time points (Day 7, 14, 35 with amoxicillin) were calculated. Isolates of these OTUs were purified for whole genome sequencing to assess the impact of amoxicillin exposure on the genomes of these particular taxa. Beta-lactam resistance genes and mutations were predicted using ARIBA and CARD (none: no beta-lactam determinants). Phylum key: B: Bacteroidetes, F: Firmicutes; P: Proteobacteria. No WGS indicates where isolates could not be recovered or purified sufficiently for genomic DNA extractions: either the glycerol stocks were contaminated or unviable. Excluded indicates OTUs that did not pass the Mash analysis and were found with 99 % identity in Day 0 metascrapes from plates containing amoxicillin. % in OTU column represents the similarity of the consensus 16S rRNA gene sequence of that OTU to reference 16S rRNA gene sequences for that species.

			Number of isolated colonies			Beta-lactam resistance genotype	
Mouse line	**OTU**	**Phylum**	**Day 0 – Amox**	**Day 0 + Amox**	**Day 7/14/35 + Amox**	**Before**	**After**
7	*Alistipes senegalensis* 98.49 %	B	1	0	1	No WGS	No WGS
7	*Bacteroides stercoris* 99.703 %	B	1	0	9	Excluded	Excluded
7	*[Clostridium] populeti* 94.435 %	F	2	0	2	None	None
7	*Coprobacillus cateniformis* 99.270 %	F	1	0	1	None	None
7	*Coprobacillus cateniformis* 92.211 %	F	6	0	2	None	None
7	*Marvinbryantia formatexigens* 92.942 %	F	0	0	8	None	None
7	*Sporosarcina newyorkensis* 99.710 %	F	9	0	1	None	None
2	*Alistipes finegoldii* 99.851 %	B	4	0	1	None	None
2	*Alistipes shahii* 99.703 %	B	1	0	17	No WGS	*cfxA*
2	*Anaerostipes caccae* 99.773 %	B	2	0	2	None	No WGS
2	*Enterococcus faecium* 99.791 %	F	19	0	48	None	None
2	*Flavonifractor plautii* 99.702 %	F	1	0	22	None	None
2	*Odoribacter splanchnicus* 99.556 %	B	2	0	2	None	*cfxA*
2	*Shigella sonnei* 99.757 %	P	2	0	7	Excluded	Excluded

To assess strain-level genomic changes such as antibiotic resistance gene acquisition, the presence of known, clinically relevant beta-lactam resistance genes and mutations in these genomes was determined using the ARIBA and CARD as in Chapter 3. Table 5.2 summarises the different isolates that were whole genome sequenced for this purpose and whether or not and they were predicted to harbour beta-lactam resistance genes. Genetic determinants of beta-lactam resistance were observed in the genomes isolated after amoxicillin treatment in the mice, but not before, in one candidate OTU (*Odoribacter splanchnicus* 99.56 %). Specifically, isolates of this OTU did not appear to contain the *cfxA* beta-lactamase before amoxicillin treatment, but did afterwards. The 198 genome sequences generated in these mouse experiments were searched for the *cfxA* gene, identifying 46 observations of this gene. All *cfxA* genes from bacteria cultured from Donor 7 mice were identical (Fig. 5.13), and all *cfxA* genes from bacteria cultured from Donor 2 mice were identical, with the only difference between the two groups being a single nucleotide polymorphism at position 775 (A in Donor 7 mouse *cfxA* genes, T in Donor 2 mouse *cfxA* genes). This suggests the *O. splanchnicus* 99.56 % isolates may have acquired the *cfxA beta-lactamase* during the course of the experiment from *Alistipes shahii* 99.703 %, *Bacteroides vulgatus* 88.703 %, or *Bacteroides ovatus* 99.48 %.

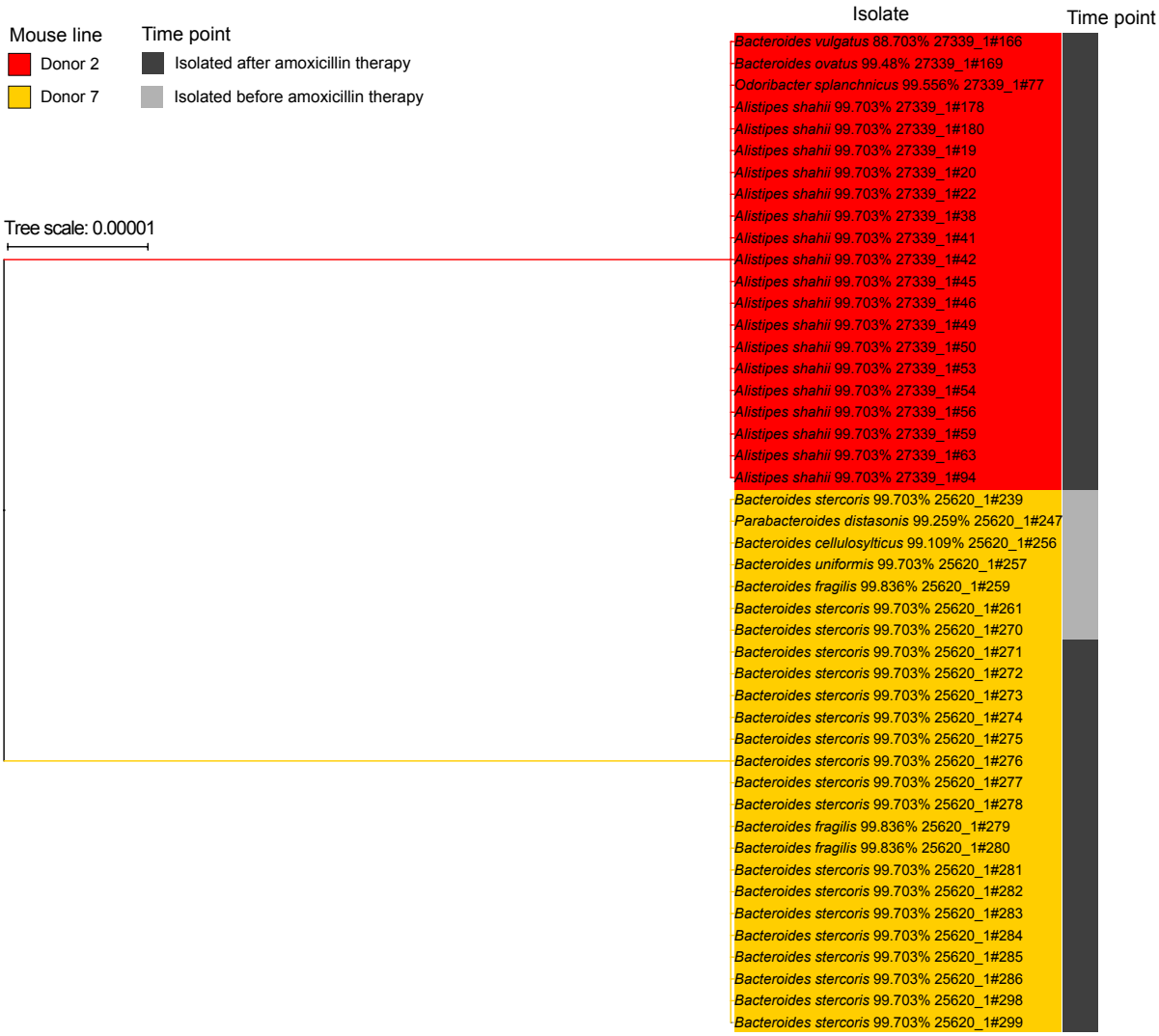

Figure 5.13. A phylogeny of the *cfxA* gene from 46 genomes from isolates cultured from mice with human-derived microbiota. Beta-lactam resistance genes and mutations were predicted using ARIBA[148] with CARD[151,152] in genomes of OTUs considered candidates of increased amoxicillin resistance (only isolated from agar containing amoxicillin after the mice received amoxicillin (Day 7/14/35) but not before (Day 0)) and other genomes generated during this study. This identified *cfxA* genes in 46 isolates; the nucleotide genes were extracted and used to infer a phylogeny, showing that *cfxA* genes were identical across genomes from a particular mouse line. No isolates with the *cfxA* gene were whole genome sequenced from the Donor 2 mouse before the mice received amoxicillin. In the Donor 7 mice, there were a number of isolates with *cfxA* sequenced from before and after the mice received amoxicillin.

The four *O. splanchnicus* isolates were cultured and three well-isolated colonies of each were used for amoxicillin MIC measurement with Etests as described previously. The two *O. splanchnicus* isolates from before amoxicillin treatment both had an average amoxicillin MIC

of 13.3 mg/L; the two isolates from after amoxicillin treatment both had an average amoxicillin of > 256 mg/L (i.e., all biological replicates of these isolates grew at the maximum amoxicillin concentration of the Etest gradient). This suggests the acquisition of the *cfxA* beta-lactamase caused an approximately 19-fold increase in amoxicillin resistance in *O. splanchnicus* 99.56 %. *O. splanchnicus* is considered strictly anaerobic and non-spore-forming[277], therefore the relative abundance of *O. splanchnicus* in the anaerobic vegetative metascrape data from plates containing amoxicillin was determined (Fig. 5.14). This species was the 30th most abundant across these samples, and the mean relative abundance increased from 0.0033 % at Day 0 to 0.022 % at the end of the experiment (Day 35). However, there were no significant differences between the mean relative abundance values of O*. splanchnicus* between consecutive time points (Day 0 vs. Day 7, Day 7 vs. Day 14, Day 14 vs. Day 35), or after either recovery periods compared to the beginning (Day 0 vs. Day 14 and 35, q values determined by unpaired t-tests with Welch's correction[276] and adjusted for multiple-testing with the Benjamini, Hochberg, and Yekutieli method[193]; q value significant < 0.05).

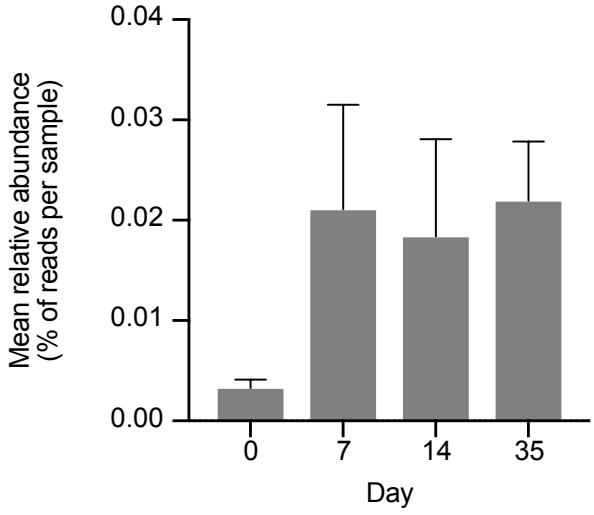

Figure 5.14. Relative abundance of *Odoribacter splanchnicus* in anaerobic vegetative metascrape samples. The relative abundance of the "strictly anaerobic, non-spore-forming" *Odoribacter splanchnicus* species was determined in the anaerobic vegetative metascrapes from plates containing amoxicillin. The relative abundance of this species increased between Day 0 and Day 7 following amoxicillin therapy in mice with human-derived microbiota, decreased slightly at Day 14 and slightly increased again by Day 35. However, these differences were not significant. Q values were determined by unpaired t-tests with Welch's correction[276], adjusted for multiple-testing with the Benjamini, Hochberg, and Yekutieli method[193]; q value significant < 0.05. Bars represent Standard Error of the Mean.

Eight of the candidate OTUs that had whole genome sequences for before and after amoxicillin therapy and were not excluded following the Mash[213] analysis did not appear to have acquired a beta-lactamase that would explain the described observations. Consequently, a comparative genomics approach as described in Chapter 4 was applied to identify potentially novel beta-lactamases. Genes in each OTU that were shared by all isolates cultured on plates containing amoxicillin after the mice were treated with amoxicillin, but absent from isolates of the same OTU cultured on plates not containing amoxicillin from before the mice received therapy, were identified. Genes fitting these criteria were searched for 90 % similarity to beta-lactamase amino acid markers as generated and described in Chapter 4. No genes were found to contain these markers with 90 % identity.

Finally, single nucleotide polymorphisms (SNPs) were searched for in the same eight OTUs as above, plus *Odoribacter splanchnicus* 99.56 %. Paired-end sequencing reads from bacteria isolated after amoxicillin treatment were mapped against assembled contigs from an isolate cultured before amoxicillin treatment. Variant bcf files were produced and used to identify SNPs in coding sequences (open reading frames (ORFs) annotated as such if over 100 bases) and non-coding regions. ORFs and non-coding regions containing SNPs were compared to known protein sequences using BLASTx or to the nr/nt database using BLASTn (Table 5.3). Twenty-two ORFs containing SNPs were most similar to enzymes involved in metabolic processes such as kinases or reductase and three ORFs were most similar to restriction endonucleases, plus one ORF was annotated as an ABC transporter ATP binding protein. However, ten ORFs were annotated as uncharacterised or hypothetical proteins.

Table 5.3. Mutations in open reading frames and non-coding regions of gut bacteria isolated after amoxicillin therapy in mice with human-derived gut microbiota. Raw sequencing reads from bacteria cultured from mice with human-derived gut microbiota after amoxicillin therapy (query samples) were mapped against reference genomes of bacteria from the same OTU that had been isolated before amoxicillin therapy. Variants were identified in open reading frames (ORFs) of more than 100 bases and non-coding regions: NS nonsynonymous; S synonymous; I insertion; SNP single nucleotide polymorphism. The sequence of ORFs and non-coding regions containing mutations were analysed with BLASTx (nr database) to identify the most similar proteins or with BLASTn (nr/nt database) to identify the most similar nucleotide sequences.

O.T.U.	Reference	Query	Feature	Mutations	Top BLAST hit
***[Clostridium] populeti* 94.435 %**	25620_1#243	25620_1#295	ORF	1 S	Sensor histidine kinase [*Eubacterium ramulus*]
			ORF	1 S	Sensor histidine kinase [*Eubacterium ramulus*]
			ORF	1 NS	TPA: beta-glucosidase [Lachnospiraceae bacterium]
			ORF	1 NS	TPA: beta-glucosidase [Lachnospiraceae bacterium]
			ORF	1 NS	Putative uncharacterized protein [*Clostridium* sp. CAG:352]
			Non-coding	1 SNP	No similarity BLASTn nr/nt or BLASTx
			ORF	3 NS	Nitroreductase family protein [*Beduini massiliensis*]
			ORF	4 NS; 1 S	Uncharacterised protein RAG0 10061 [*Rhynchosporium agropyri*]
			ORF	1 S	TPA: phosphoenolpyruvate carboxykinase (ATP) [*Blautia* sp.]
			ORF	1 NS	Phosphoenolpyruvate carboxykinase [ATP] [[*Clostridium*] *clostridioforme* 90A7]
			ORF	1 S	MULTISPECIES: phosphoenolpyruvate carboxykinase (ATP) [Clostridiales]
			ORF	1 S	ABC transporter ATP-binding protein [*Flavonifractor plautii*]
			Non-coding	1 SNP	No similarity BLASTn nr/nt or BLASTx
			ORF	5 NS; 2 S	Gfo/Idh/MocA family oxidoreductase [*Anaerobium acetethylicum*]
			ORF	6 NS; 1 S	Gfo/Idh/MocA family oxidoreductase [*Anaerobium acetethylicum*]
			ORF	1 NS	Glucan 1,6-alpha-glucosidase [*Clostridium* sp. KLE 1755]
			ORF	1 NS	Alpha-glucosidase [*Eisenbergiella massiliensis*]
***Coprobacillus cateniformis* 99.270 %**	25620_1#240	25620_1#297	Non-coding	1 SNP	Helix-turn-helix domain-containing protein [*Massilimicrobiota timonensis*]
***Coprobacillus cateniformis* 99.211 %**	25620_1#293	25620_1#300	ORF	1 S	Hypothetical protein [*Coprobacillus cateniformis*]
***Marvinbryantia formatexigens* 92.942 %**	25620_1#287	25620_1#309	ORF	1 NS	Peptidase [*Anaerotruncus* sp. 1XD22-93]

O.T.U.	Reference	Query	Feature	Mutations	Top BLAST hit
			ORF	1 NS	Peptidase [Lachnospiraceae bacterium] or Penicillin-binding protein A [uncultured *Clostridium* sp.]
***Sporosarcina newyorkensis* 99.710 %**	25620_1#231	25620_1#325	ORF	1 NS	Preprotein translocase subunit SecA [*Sporosarcina newyorkensis*]
			ORF	1 NS; 1 S	TrkH family potassium uptake protein [*Sporosarcina newyorkensis*]
***Sporosarcina newyorkensis* 99.710 %**	25620_1#232	25620_1#325	ORF	1 NS; 1 S	TrkH family potassium uptake protein [*Sporosarcina newyorkensis*]
***Alistipes finegoldii* 99.851 %**	27261_7#7	27339_1#174	ORF	1 NS	Hypothetical protein/restriction endonuclease
			ORF	1 NS	Restriction endonuclease
			ORF	1 I	Leucine-rich repeat domain-containing protein [*Alistipes finegoldii*]
			ORF	1 I	Leucine-rich repeat domain-containing protein [*Alistipes finegoldii*]
			ORF	1 I	Leucine-rich repeat domain-containing protein [*Alistipes finegoldii*]
			ORF	1 I	Uncharacterised protein BN754_01505 [*Alistipes finegoldii* CAG:68]
			ORF	2 I	Glycosyl transferase
			ORF	3 NS; 5 S	Restriction endonuclease
			ORF	2 NS: 5 S	Restriction endonuclease
***Enterococcus faecium* 99.791 %**	27339_1#100	27339_1#138	ORF	1 NS	Heavy metal translocating P-type ATPase [*Enterococcus faecium*]
			ORF	1 NS	MULTISPECIES: aldo/keto reductase [*Enterococcus*]
***Flavonifractor plautii* 99.702 %**	27339_1#70	27339_1#26	ORF	1 NS	No similarity BLASTx; BLASTn similarity to *Flavonifractor* genome
			ORF	1 NS	Acetaldehyde dehydrogenase [*Flavonifractor plautii* ATCC 29863]
			ORF	1 S	Acetaldehyde dehydrogenase [*Flavonifractor plautii* ATCC 29863]

5.3 Discussion

In this chapter I have studied the impact of amoxicillin therapy on amoxicillin resistance in mice with human-derived gut microbiota. Reference genome based metagenomics was used to observe changes to the pre-existing amoxicillin-resistant community that lasted beyond the end of the antibiotic treatment. A recent study on amoxicillin combined with clavulanate therapy in healthy adult humans also found gut microbiota compositional changes immediately after one week's treatment[278]. Two weeks after treatment stopped, these changes were no longer apparent and the gut microbiota communities had reverted back to pre-treatment profiles[278]. However, that study mainly used 16S rRNA sequencing for profiling the gut community and only looked at relative abundances of bacterial families, whereas I have assessed the species level. In addition, the combination of culturing and WGS to generate study-specific reference genome databases for taxonomic classifications means my analysis in this thesis is tailored to the gut microbiomes being studied, providing higher taxonomic resolution. This may explain the disparity between the two studies.

The amoxicillin-resistant community was studied in mice with human-derived microbiota by culturing faecal samples on agar plates, including plates containing amoxicillin at 8 mg/L amoxicillin, a concentration above which pathogenic bacteria can be considered resistant to amoxicillin. This was further divided into the aerobic, anaerobic vegetative and anaerobic spore-forming communities. The key findings from these humanised microbiota mouse experiments are illustrated in Figure 5.15. The amoxicillin therapy appeared to have an impact on the bacterial load of each of these communities, including the reduction of the amount of anaerobic spore-forming bacteria. An overall reduction in the total amount of gut bacteria following amoxicillin has been observed in previous studies[179]. However, in this study

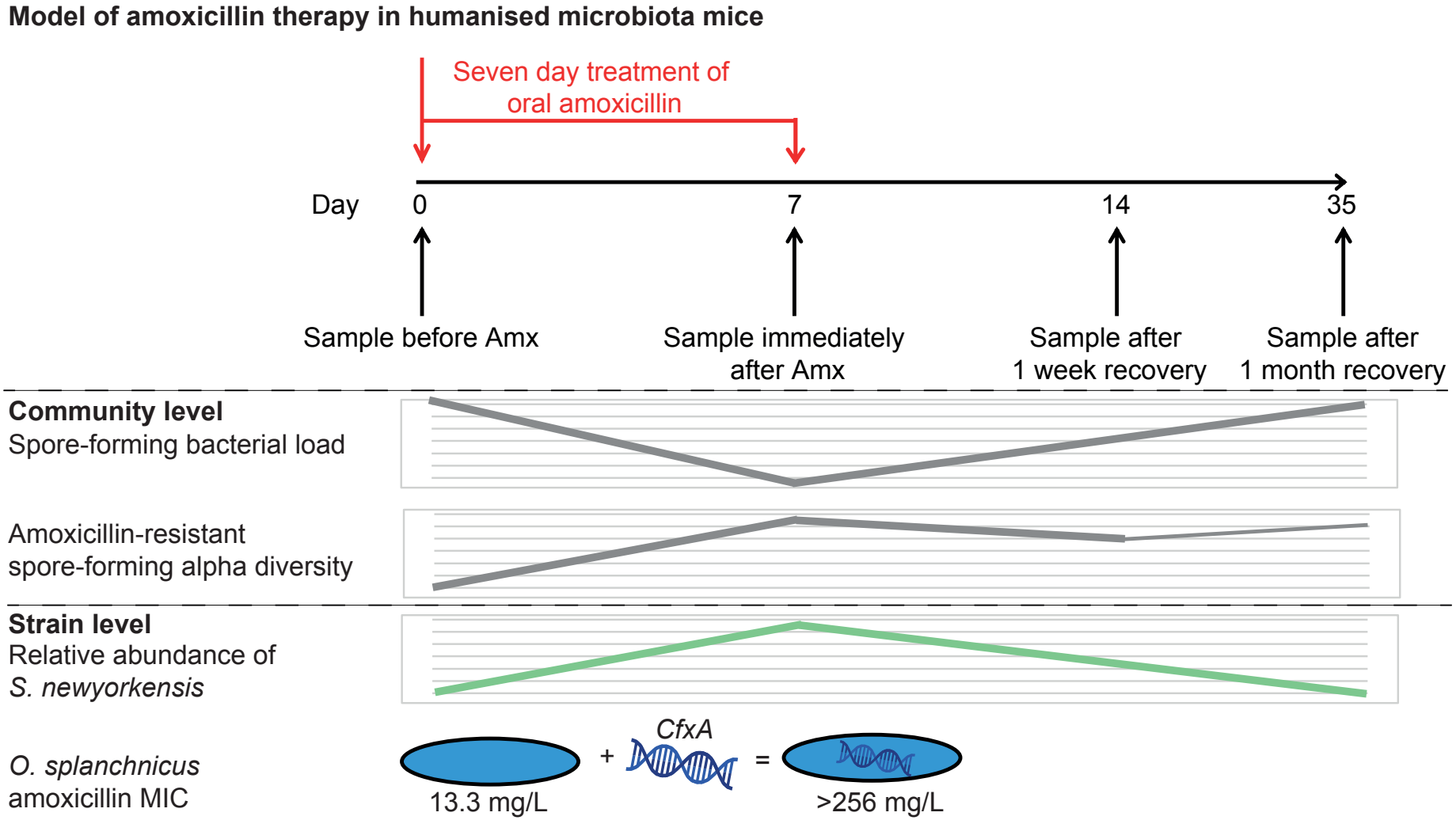

Figure 5.15. The key findings from the humanised microbiota mouse experiments. An illustration of the changes to the amoxicillin-resistant spore-forming community (bacterial load and alpha diversity) and individual strains (relative abundance of *Sporosarcina newyorkensis* and acquisition of *cfxA* by *Odoribacter splanchnicus*). Mice with humanised gut microbiota were given amoxicillin therapy for seven days at a therapeutic concentration. Faecal pellets were collected before and after therapy to allow culturing of individual isolates, WGS and whole genome shotgun metagenomic sequencing of metascrape samples.

there was a slight increase in anaerobic vegetative bacteria – perhaps utilising the space that had been made available by reductions in other bacteria. The amoxicillin-resistant anaerobic spore-forming bacteria also showed significant reductions in load; this indicates that the recommended therapeutic dose of amoxicillin[206] is sufficient to kill some bacteria considered resistant. This is in line with what is expected for amoxicillin – even though it is considered broad spectrum, it targets more Gram-negatives than Gram-positives[31].

That spore-formers, a major but comparatively understudied part of the gut microbiome, are significantly reduced by amoxicillin highlights that the impact of antibiotics on the gut microbiota is not yet fully understood; this result is unexpected given the described amoxicillin spectrum of activity. To my knowledge, this is the first time the spore-forming community has specifically been studied for its response to antibiotics, by using a combination of culturing, WGS and reference genome based metagenomics. Antibiotic treatment can promote spore formation[279] and the assumption could be made that spore-forming bacteria would therefore be protected against significant effects to their total abundance in the gut. However, if the spores re-enter the vegetative state they may not be protected and be impacted by the amoxicillin treatment. This could result in fewer spore-forming bacteria present to produce spores; during the targeted culturing there would then be fewer spores present. After the amoxicillin therapy ends, the spore-formers are able to recover and produce spores again, hypothetically leading to the findings observed in this chapter. In the future, quantitative methods for measuring bacterial load may provide more precise and insightful information[280].

Comparing alpha diversity in the amoxicillin-resistant spore-formers appeared to suggest that the richness of this community increased following amoxicillin treatment. This could be explained by previously susceptible species developing resistance. However, the amoxicillin-resistant anaerobic vegetative and spore-forming communities comprised mostly lowly

abundant or rare species (> 90 % cumulative relative abundance), compared to the amoxicillin-resistant aerobic community. The increase in richness could be due to a reduction in species such as *E. faecalis*, allowing lowly-abundant ones such as *S. newyorkensis* to reach the detectable limit in the metagenomic sequencing analyses and appear more relatively abundant.

The amoxicillin-resistant aerobic community was initially dominated by *E. faecalis*, a Gram-positive facultative anaerobe that is common in the gut[281]. This species was significantly less relatively abundant following amoxicillin therapy. In previous studies of antibiotics on the gut microbiome, enterococci have been observed to expand in the gut following antibiotic therapy[282]; however, this is dependent on the type of antibiotic. Amoxicillin has previously been shown to reduce Gram-positive cocci such as enterococci and increase enterobacteria instead[283,284]. Here, there was a small but insignificant increase in the relative abundance Enterobacteriaceae member *E. coli*. Two *Bacteroides* species (*B. fragilis* and *B. cellulosilyticus*) were dominant at Day 14, and three other *Bacteroides* species (*B. thetaiotaomicron*, *B. uniformis* and *B. vulgatus*) were also in the top ten most relatively abundant species at this time point in aerobic metascrapes. This is surprising given *Bacteroides* species are thought to be obligate anaerobes[148]. Colonies of these five *Bacteroides* species plus *B. faecis*, *B. stercoris* and *B. xylanisolvens* were all isolated from aerobic culture plates, indicating that this was not a "false positive" result in the metagenomic data. *B. fragilis* has been reported as growing in microaerophilic environments (≤ 0.05 % oxygen)[285], and specific DNA mutations have resulted in strains being able to grow in up to 2 % oxygen. In addition, *B. thetaiotaomicron* has been shown to express scavenging enzymes when exposed to oxygen to degrade oxygen compounds and reduce their toxicity[286]. The *Bacteroides* strains growing under aerobic conditions here may have contained or developed mutations that were selected for, allowing

their growth under aerobic conditions. Whether or not this was caused by the amoxicillin treatment or by exposure to oxygen is not determinable in this study. How *Bacteroides* might adapt to aerobic conditions would be interesting to investigate in the future, as this would provide novel insights into their biology.

The most obvious increase in relative abundance in the aerobic metascrapes was that of *S. newyorkensis*. *S. newyorkensis* is another Gram-positive cocci but one that is rare in its ability to form spores; it is also facultatively anaerobic. This may explain its presence and increase in relative abundance in the amoxicillin-resistant aerobic community (Fig. 5.15), following the reduction of *E. faecalis*. *Sporosarcina* species have only recently begun to be studied using genomics[287], with *S. newyorkensis* first being described in 2012[288]. This species was first isolated from human blood[288], which is typically considered a sterile environment and as such, the presence of bacteria in blood indicates infection. Therefore, this study provides insights into the effect of amoxicillin on newly-discovered bacteria, including ones that may have clinical relevance through the potential to cause infections. Another recent study has identified an increase in the relative abundance of a spore-former following beta-lactam treatment: *Candidatus* Borkfalkia ceftriaxoniphila[289] was relatively lowly abundant in the gut prior to ceftriaxone treatment in humans and then dominated the community after ceftriaxone treatment[264]. There, *Candidatus* Borkfalkia ceftriaxoniphila was considered a potential keystone species in the recovery of the gut community following antibiotics; perhaps these temporarily dominant spore-formers aid the re-establishment of a diverse gut microbiome.

In addition to examining the impact of amoxicillin on the amoxicillin-resistant community, I studied the impact on individual strains of gut bacteria. Detecting changes in the genotypic and phenotypic resistance profiles before and after therapeutic antibiotic exposure using

broad identified the acquisition of a known beta-lactamase by a commensal strain of gut bacteria). The *cfxA* beta-lactamase gene is localised on a mobile transposon[290], and since being first described in *B. fragilis*[290], *cfxA* and related genes have also commonly been found in *Prevotella*, *Porphyromonas* and *Capnocytophaga* species[291]. *Prevotella* species are common members of the gut microbiota[292] and *cfxA* in *Prevotella* in particular is associated with increased beta-lactam resistance in periodontal infections[292]. Thus, the presence of *cfxA* in gut microbiota that can act as opportunistic pathogens has potential clinical relevance.

Here, *cfxA* appears to have been acquired by an *O. splanchnicus* strain following amoxicillin therapy, possibly from another Bacteroidetes species, apparently conferring a 19-fold increase in amoxicillin resistance (Fig. 5.15). *O. splanchnicus* has been described as an opportunistic pathogen[277]; if we extrapolate beyond these experiments, the ability of *O. splanchnicus* to acquire beta-lactamases from the gut microbiome therefore has direct potential clinical relevance. A recent porcine model of amoxicillin treatment showed that amoxicillin promoted the enrichment of beta-lactamases including *cfxA*[293]. However, as previously discussed, using whole genome shotgun metagenomics to assess gene prevalence does not allow for the hosts of those genes to be identified, nor can it distinguish whether an increase in gene abundance is caused by the spread of that gene to additional species or the relative increase in abundance of the original host species. Here, I have used broad culturing and WGS to discriminate new hosts of antibiotic resistance genes. Therefore, the unique combination of humanised gnotobiotic mouse models, advanced culturing and genomic analysis in this chapter provides novel and deep understanding of how antibiotic treatment changes the resistance profile of the gut microbiota community and individual strains, with implications relevant to the dissemination of antibiotic resistance and the evolution of drug-resistant infections.

There were other candidate OTUs that appeared to have acquired amoxicillin resistance over the course of the experiment in the absence of known beta-lactam resistance genes or mutations using the methods applied in this study. These OTUs could represent pre-existing amoxicillin-resistant species that have increased in relative abundance above the detection level. Alternatively, it could be due to other resistance mechanisms, such as developing mutations not listed in CARD that increased resistance. This would be in line with the issues predicting resistance in commensal gut bacteria using genomics from the previous chapter, and further highlights how improved annotation of intestinal microbiota would be beneficial. Isolates of these OTUs from after the amoxicillin treatment had SNPs compared to isolates from before the antibiotic therapy. The majority of the ORFs containing variant sites were involved with metabolic processes. One was annotated as a transporter protein, which could potentially be involved in the efflux or export of amoxicillin. In addition, one was similar to a penicillin-binding protein; the observed non-synonymous mutation could have resulted in a reduced affinity for amoxicillin. Several of the ORFs were annotated as hypothetical or uncharacterised proteins; thus their function is unclear. Potentially these may also be candidate novel amoxicillin-resistance genes – though this additionally emphasises the issues surrounding annotation of bacterial genomes and proteins.

Overall, this chapter shows that amoxicillin has significant impacts on the amoxicillin-resistant community of gut bacteria in mice with human-derived microbiota, including relatively uncharacterised spore-forming bacteria. Moreover, a commensal Bacteroidetes strain of *O. splanchnicus* acquired the *cfxA* beta-lactamase gene from the surrounding gut microbiome following exposure to amoxicillin *in vivo* in these mice. Thus, this chapter further emphasises the role of the human gut microbiome as a reservoir for antibiotic resistance in terms of its antibiotic-resistant community- and strain-level changes following antibiotic treatment.