Chapter 3: Characterisation of genomic antibiotic resistance in commensal gut bacteria

3.1 Introduction

The gut microbiome is considered a reservoir of antibiotic resistance[214] through the sharing of antibiotic resistance genes (ARGs) amongst autochthonous, commensal gut bacteria. In addition, these ARGs may be shared with allochthonous, transient bacteria passing through the gastrointestinal tract, which can include bacterial pathogens[123]. This is evidenced by studies that have identified both known, clinically relevant antibiotic resistance genes in samples collected from the gut microbiome as well as 'novel' antibiotic resistance genes (not previously identified or seen in pathogenic bacteria). For example, the functional metagenomics study by Sommer, Dantas and Church in 2009[139] identified ARGs identical at the nucleotide level to genes previously identified in clinical isolates of disease-causing bacteria, proving that the gut microbiota do harbour clinically relevant ARGs. However, the majority of antibiotic resistance-conferring inserts (82 %) were distantly related to known antibiotic resistance genes from pathogens (nucleotide identity 60.7 % on average)[139]. These results highlight both the diversity of genetic antibiotic resistance determinants in the gut but also their potential to be shared between commensal and pathogenic bacteria.

Another interesting study of the gut as a reservoir of antibiotic resistance used whole genome shotgun metagenomic sequencing to examine the resistomes of 180 healthy individuals from 11 different countries across Europe, Asia, North and South America representing industrialised, low-income and remote societies[215]. In total, 507 different ARGs were identified, including eight shared by all 180 individuals[215]. This further highlights how antibiotic resistance appears to be widely distributed across gut microbiomes. If those ARGs

are acquired by pathogenic bacteria passing through the gut, or if the gut bacterium hosting the ARG moves outside of its usual location in the gut and causes an infection elsewhere in the body, then gut microbiota are part of the antibiotic resistance problem.

These types of metagenomic studies investigate the entire community but not individual bacteria, therefore it is essentially impossible to say exactly which bacterium a particular antibiotic resistance determinant might have come from. Antibiotic resistance genes are often located on mobile genetic elements; since mobile genetic elements can be shared between different bacteria, it is hard to place the nucleotide sequence of a mobile genetic element identified from a mixed sample into its original genomic context. It can also be difficult to validate the results, as without knowing the host or having an isolate of the suspected host a phenotype cannot be measured and correlated with the presence or absence of the ARG in question. Therefore, these studies are good for observing the presence and abundance of known or putative ARGs in an environment, but not determining exactly which bacteria they belong to. This means that we do not necessarily fully understand the taxonomic placement of antibiotic resistance determinants among the bacterial community in a mixed sample, such as a stool sample that represents the gut microbiome. Combined with the fact that we are still discovering new members of the gut microbiota through metagenome-assembled genomes and high-throughput culturing[146,151,216] the full potential of the gut microbiome as an antibiotic resistance reservoir remains to be understood.

With recent developments in culturing of anaerobic commensal gut bacteria[146,151], there is now the opportunity to use these methods to determine a comprehensive map of antibiotic resistance determinants in these diverse and relatively uncharacterised microorganisms. It is important to determine which commensal gut bacteria harbour antibiotic resistance genes as this information will help understand which ones are of concern for acting as donors to spread

antibiotic resistance. In this chapter, I use a unique collection of reference genomes to study the presence of known antibiotic resistance determinants in diverse commensal human gut bacteria.

3.2 Results

3.2.1 Summary of resources used in this chapter

To determine the presence of antibiotic-resistant determinants in diverse commensal human gut bacteria, I have used the genomes from the Human Gastrointestinal Bacteria Culture Collection (HBC)[151]. Each isolate has been whole genome sequenced using Illumina short-read paired-end sequencing, assembled and annotated[183], and the physical isolates are held in glycerol stocks at -80 °C allowing for phenotypic validation and characterisation (Chapter 4). The HBC contains 737 gut bacteria isolated from healthy adult humans, who had not taken antibiotics in the six months prior to sampling, using broad range culturing and targeted phenotype culturing for spore-forming bacteria[146,151] (Fig. 3.1). The collection contains 273 species in total, 105 of which are considered novel (Table 3.1).

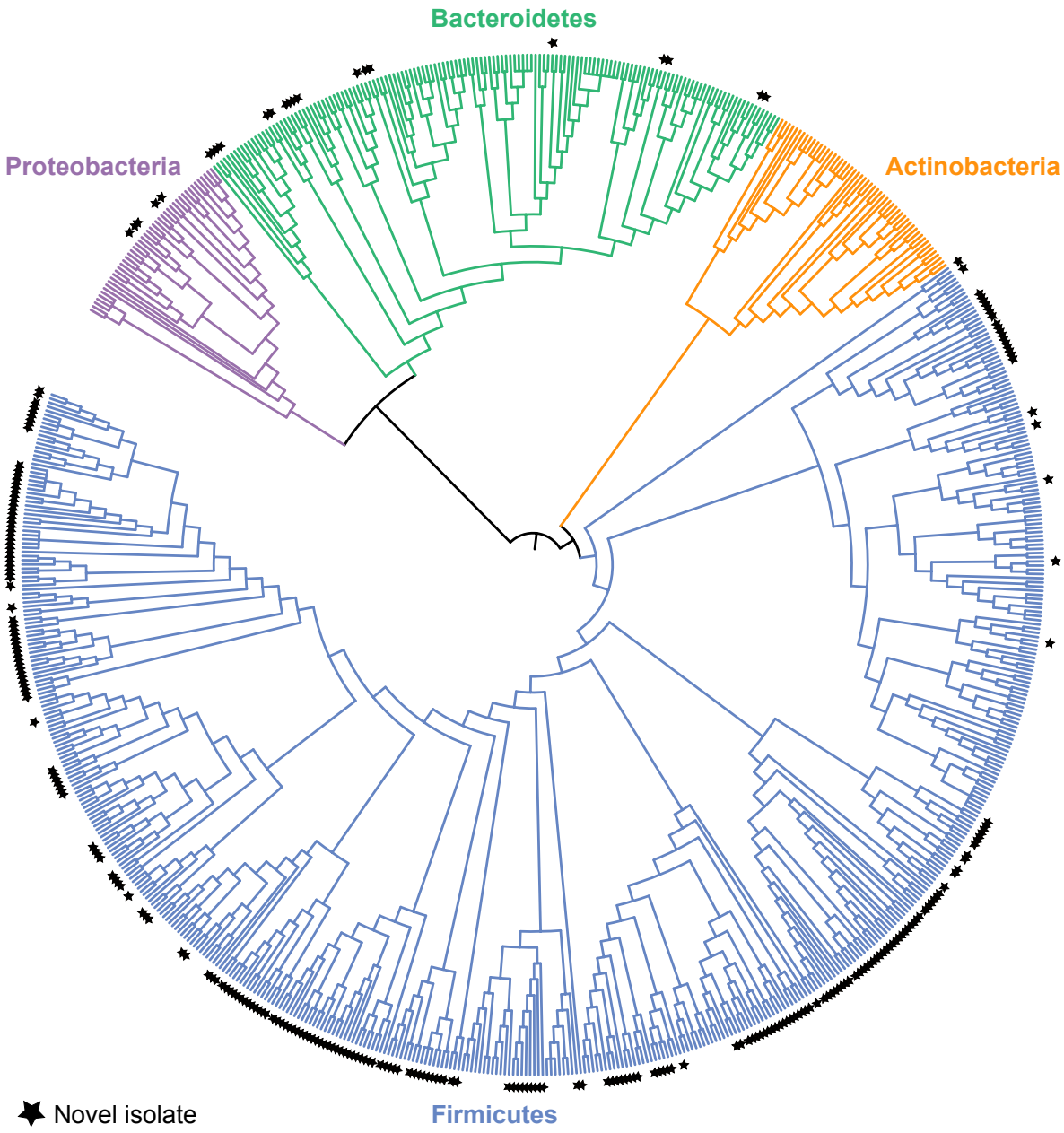

Figure 3.1 Phylogeny of the HBC commensal gut bacteria. The Human Gut Bacteria Culture Collection contains 737 isolates of human gut bacteria[151]. The amino acid sequences of 40 core genes were extracted from these genomes and used to infer a phylogeny to illustrate the taxonomic diversity of this culture collection. 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.

Table 3.1. Taxonomic information for the HBC. The number of total or novel isolates, species and families is summarized for the four phyla of the HBC and the HBC overall.

	Actinobacteria	Bacteroidetes	Firmicutes	Proteobacteria	All HBC
Total isolates	53	143	496	45	**737**
Novel isolates	0	18	253	5	**276**
Total species	16	40	203	14	**273**
Novel species	0	13	91	1	**105**
Total families	6	4	18	3	**31**
Novel families	0	1	0	0	**1**

Species were defined by considering the sequence identity of each genome's full length 16S rRNA gene. 16S rRNA gene sequences at least 97.8 % identical to each other are considered the same species[182]. In addition, the HBC 16S rRNA gene sequences were compared to RefSeq 16S rRNA sequences to assign taxonomic classification. Any 16S rRNA gene sequences of less than 94.5 % similarity to known sequences are considered to belong to novel genera, < 86.5 % to novel families, < 82.0 % novel orders, < 78.5 % novel classes and 75.0 % novel phyla[182]. In total, there are 276 isolates that belong to novel taxonomic groups in the HBC. This unique collection offers extensive and novel phylogenetic diversity, compared to the six ESKAPE pathogenic species (*E. faecium, S. aureus*, *K. pneumoniae*, *A. baumannii*, *P. aeruginosa* and *Enterobacter* spp.) that represent just six species from two orders within the Firmicutes and Proteobacteria. ESKAPE pathogens are included later as a comparator as they have been the focus of most antibiotic resistance gene and mutation research.

Sequence reads for each genome were searched against the Comprehensive Antibiotic Resistance Database (CARD)[163,164] as it is one of the largest resistance determinant databases, contains both ARGs and resistance-associated mutations, and is updated regularly. It is also commonly used as a reference database for studying antibiotic resistance genes and

mutations (e.g.[143,215,217,218]). Version 2.0.2 of CARD (from June 2018, analysis performed in July 2018), contains 2426 antibiotic resistance determinants. The CARD ontology describes which antibiotics each determinant confers resistance to; these descriptions were further grouped by the major antibiotic class – for example all sub-classes of beta-lactams were combined (Fig. 3.2; see Appendix 1 for Table A1.1 describing groupings). In total, there are 29 different antibiotic classes or categories represented in CARD. Determinants conferring resistance to Macrolide, Lincosamide, Pleuromutilin and Streptogramin antibiotics were grouped together (MLPS) as resistance determinants against these antibiotics can have cross-resistance each other[219] and are grouped together in a single, separate category. Any other determinants described as conferring resistance to more than one class of antibiotics were grouped under "non-specific antibiotic resistance" for the purpose of this study.

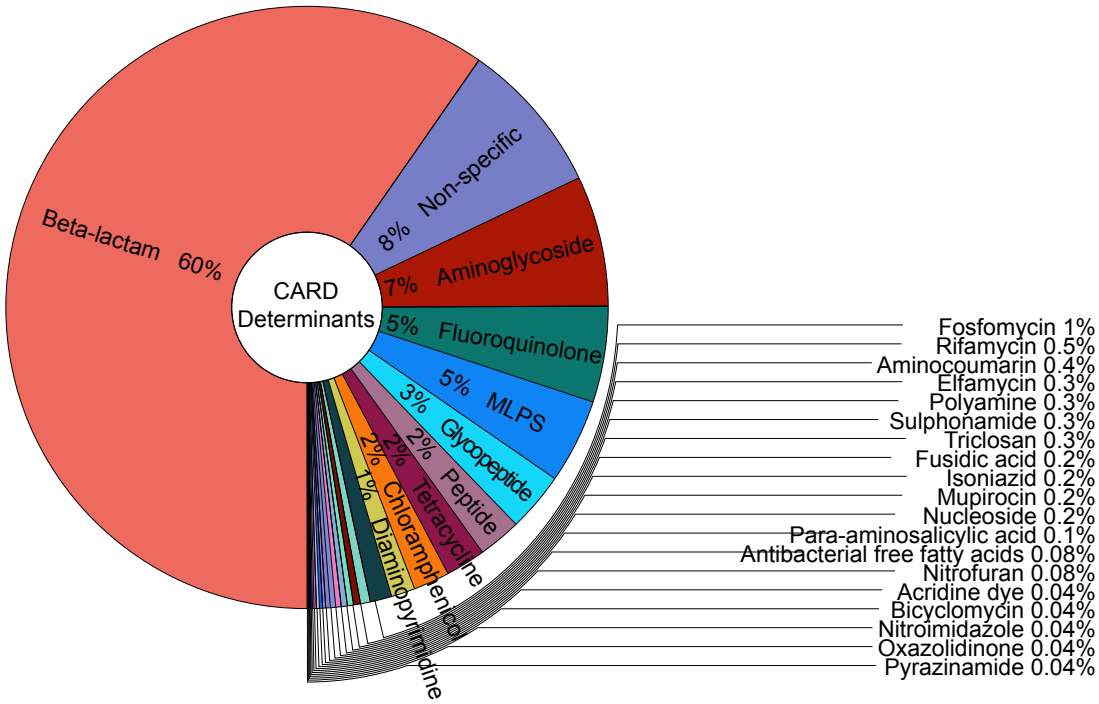

Figure 3.2. The proportions of antibiotic categories in the Comprehensive Antibiotic Resistance Database. 2426 antibiotic resistance determinants are described in CARD. The CARD ontology describes which antibiotics each determinant confers resistance to. If determinants were described as conferring resistance to multiple classes of antibiotics, these were grouped under "non-specific" 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.

3.2.2 Computational predictions of antibiotic resistance in 737 whole genome sequences of anaerobic gut bacteria

Approximately 60 % of the determinants in CARD are associated with beta-lactam resistance (Fig. 3.2). The CARD database includes redundancy, therefore I used ARIBA[160] (Antibiotic Resistance Identification By Assembly) to predict the presence of CARD determinants in the HBC genomes. ARIBA clusters gene sequences from CARD by similarity, so that only the presence of "unique" non-redundant antibiotic resistance determinants are considered. Following the ARIBA 'prepareref' command, which performs the clustering of determinants, 1024 clusters were created. ARIBA then performs local assemblies of sequence reads, in this case for each individual genome, against the reference sequence for each antibiotic resistance determinant cluster. If an assembled gene was 90 % similar to a reference antibiotic resistance determinant cluster sequence at the nucleotide level, that cluster was reported as present in that genome.

In total, 178 unique clusters of genetic antibiotic resistance determinants were observed in the HBC (Fig. 3.3, see Appendix 2, Table A2.1 for full description and https://docs.google.com/spreadsheets/d/1zwmhUicOW3JVW_9y6P6LssbavW47EFiRq4_wwnS9CMg/edit?usp=sharing for Table A2.2), with a combined total of 1434 occurrences across the 737 genomes. The determinants were grouped as described before: 19 of the 29 possible categories of antibiotics were represented by the HBC genomic resistances. The largest proportion of identified resistance determinants (29 %) were those considered "non-specific" determinants (Fig. 3.3). Beta-lactam antibiotic resistance determinants were the second most common type (13 %) to be observed. All antibiotic classes from the WHO List of Essential Medicines[21] are represented in these observations, with the exception of oxazolidinone; it is

possible oxazolidinone resistance may be present via non-specific determinants if they alter the 23S and 30S rRNA subunits targeted by this antibiotic.

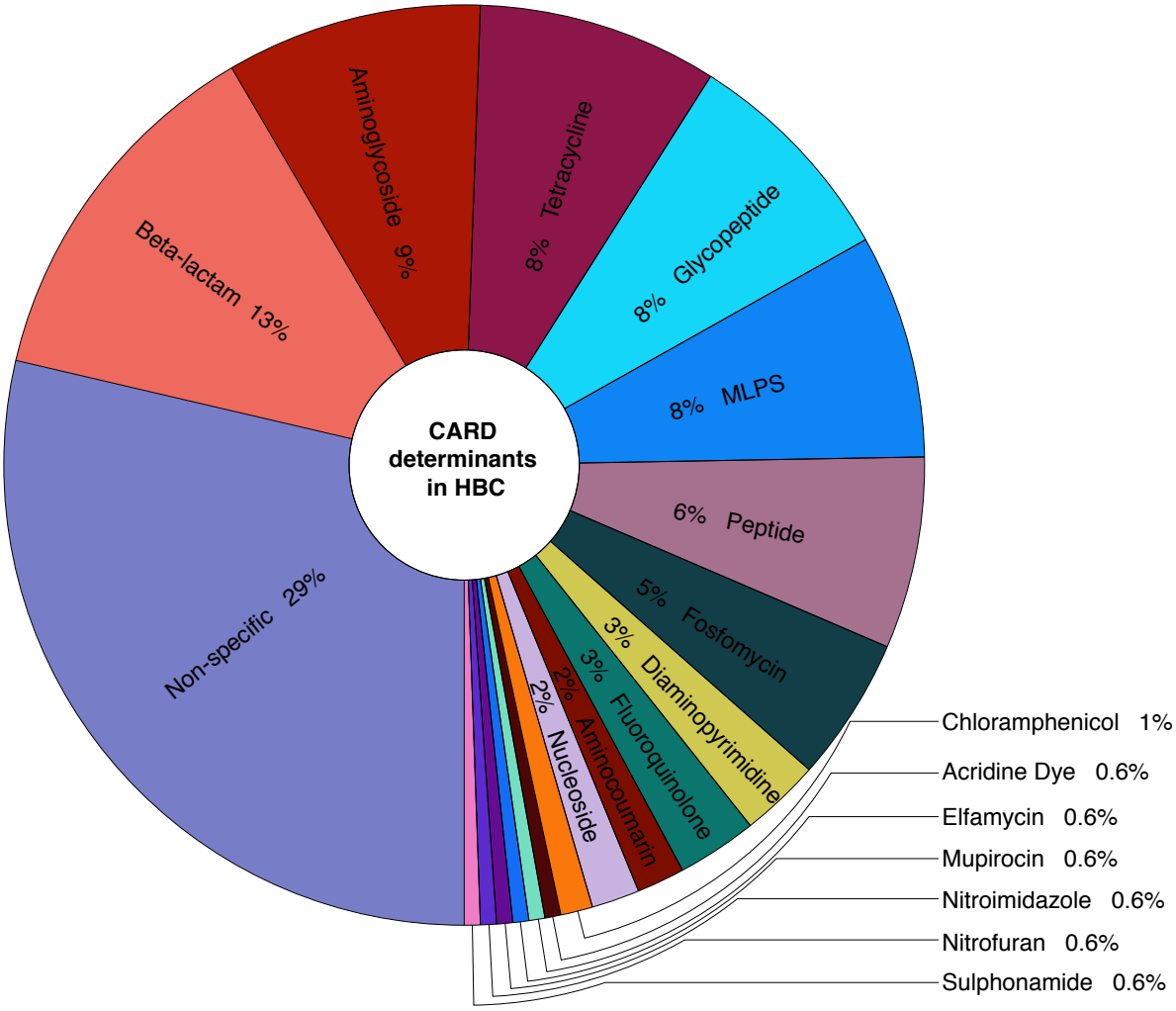

Figure 3.3. Proportions of 178 antibiotic resistance determinants identified in 737 isolates of human gut bacteria. The determinants were grouped by the antibiotic class that they are described as conferring resistance to according to the CARD ontology. MLPS = Macrolide, Lincosamide, Pleuromutillin and Streptogramin A.

Approximately half of the HBC genomes (390/737; 52.9 %) were predicted to contain at least one antibiotic resistance determinant (Fig. 3.4A). The number of unique determinants in a single genome ranged from 1 to 56 (including non-specific resistance determinants; Fig. 3.4B), although approximately three-quarters of the genomes were observed to contain just one or two antibiotic resistance determinants. Individual isolates carried resistances to 10 different classes of antibiotics (excluding the non-specific antibiotic resistance category; Fig. 3.4C),

though again approximately three quarters of the HBC only harboured resistance to two or fewer antibiotic categories. Overall, the majority of this phylogenetically diverse collection of human commensal gut microbiota contain antibiotic resistance determinants, despite not having been exposed to antibiotics for at least six months. In addition, the range in number of resistance determinants, and the number of antibiotic classes resistance is predicted to, per isolate indicates variability in antibiotic resistance genotypes across the HBC.

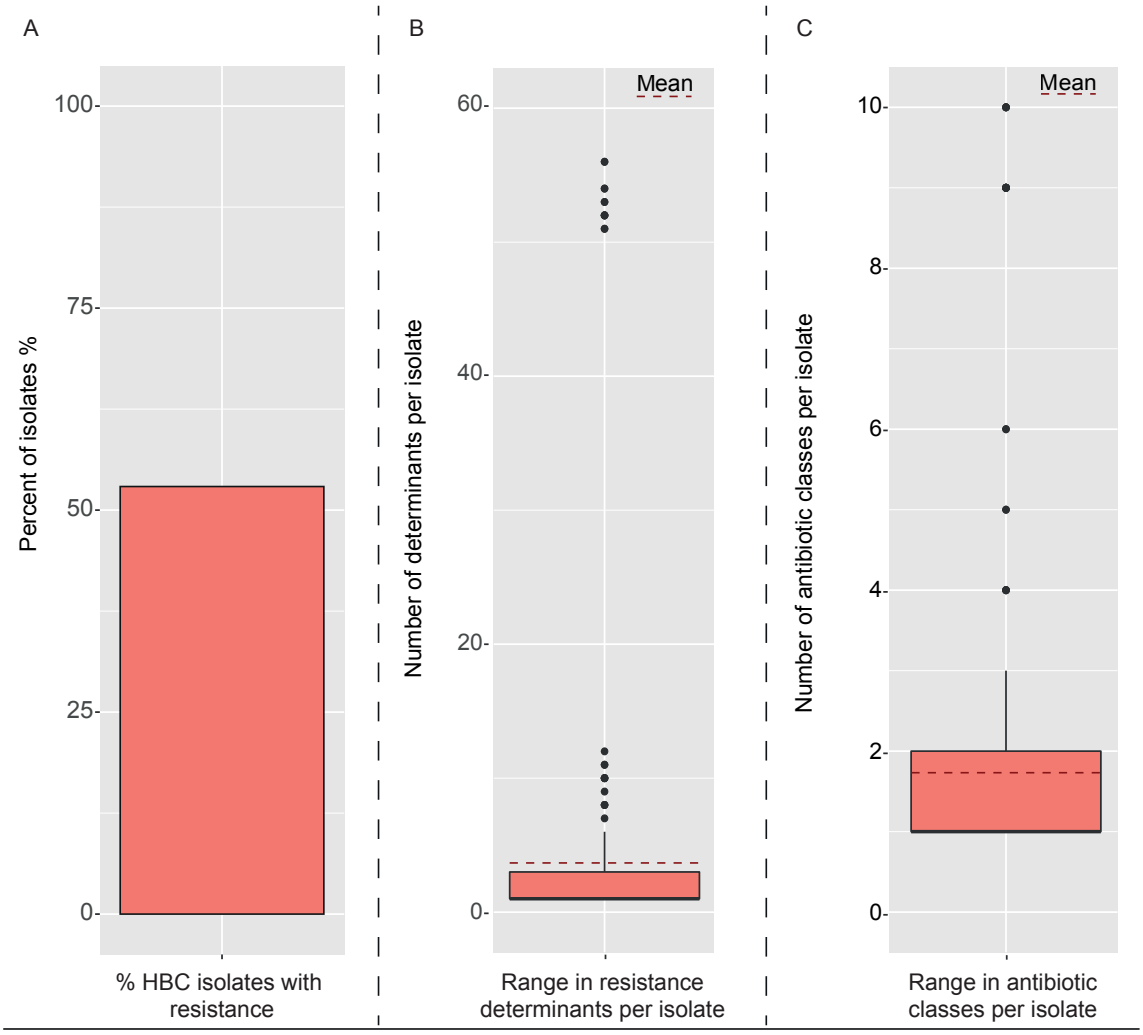

Figure 3.4. Summary of genetic determinants of antibiotic resistance in the HBC.
A) The proportion of the 737 genomes with genetic determinants of resistance identified using CARD and ARIBA.
B) The range in the number of genetic determinants of resistance predicted in an individual genome.
C) The range in different antibiotic classes an individual genome was predicted to be resistant against (excluding non-specific antibiotic resistance).
For B) and C), the boxes show the interquartile range determined using the Tukey method; the black circles indicate outlier observations and the thick black line represents the median. The red dashed line represents the mean.

3.2.3 Variation of predicted genomic resistance across the four key gut bacteria phyla

The presence of a resistance determinant to a particular antibiotic is assumed to confer phenotypic resistance to that antibiotic (e.g., the presence of a beta-lactam resistance gene predicts that isolate to be resistant to beta-lactam antibiotics). Therefore in this thesis these observations are considered "predicted resistance". The distribution of predicted resistances belonging to the 19 categories of antibiotic among the 737 genomes is demonstrated in Figure 3.5. This shows that antibiotic resistance is predicted throughout all four phyla in the HBC, and that predicted resistances vary, even between closely related isolates. However, Proteobacteria, and certain members of the Firmicutes, have more predicted antibiotic resistances than other isolates. These isolates belong to species known to be able to act as opportunistic pathogens such as *Enterobacter cloacae*[220], *Klebsiella oxytoca*[221], *K. pneumoniae*[221] (Proteobacteria) and *E. faecalis*[222] and *E. faecium*[222] (Firmicutes). These initial findings support the role of the gut microbiota as a reservoir of antibiotic resistance and its potential to contribute to antibiotic-resistant infections.

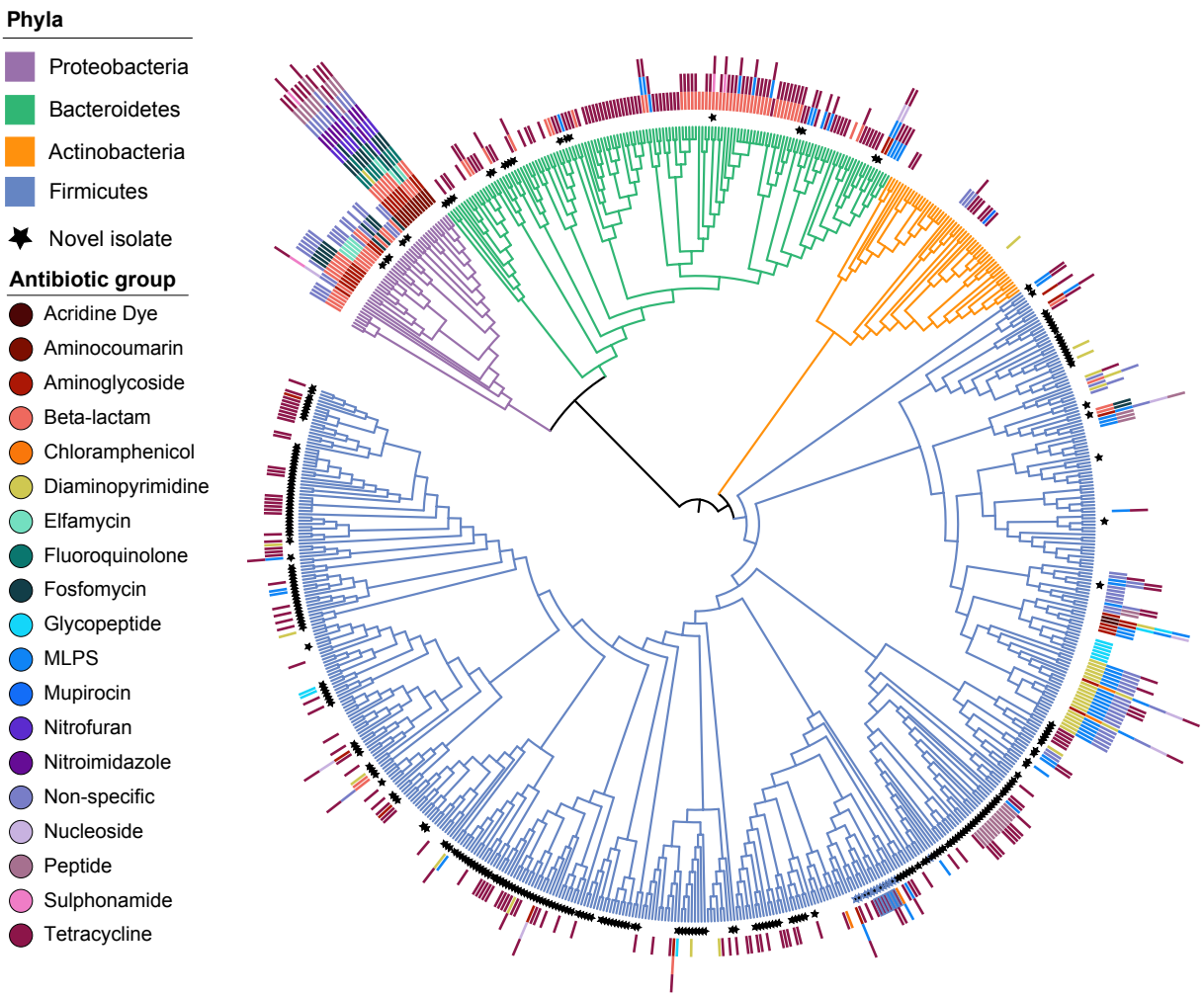

Figure 3.5. Observations of predicted antibiotic resistance in the HBC isolates against the core genome phylogeny. The phylogeny represents the core genomes of 737 whole genome sequences of gut microbiota isolated from healthy human faecal samples. Antibiotic resistance genes and mutations described in CARD were identified in these genomes, which were grouped by the corresponding class of antibiotic. If more than one, then these were classified as 'non-specific resistances' or MLPS if Macrolide, Lincosamide, Pleuromutillin or Streptogramin A. The outer rings of coloured bars show the presence of at least one resistance determinant to a particular antibiotic class. Proteobacteria isolates appear to have the highest number of resistances to different antibiotic classes.

Having determined the overall occurrence of known antibiotic resistance determinants in the HBC and gained a broad idea of their distribution, I next sought to understand the prevalence of antibiotic resistance predicted in each of the four main gut microbiota. From the phylogeny in Figure 3.5, Proteobacteria have the highest number of resistances to different antibiotic classes. Looking more specifically at the proportion of isolates within phyla (Fig. 3.6) further demonstrates this: 95.6 % of Proteobacteria are predicted to harbour resistances compared

to 82.5 % of Bacteroidetes, 43.5 % of Firmicutes and just 24.5 % of Actinobacteria. Proteobacteria and Bacteroidetes also have a significantly higher proportion of predicted resistant isolates than expected based on the HBC proportion overall (q values of < 0.0001, determined by Fisher exact tests and corrected for multiple testing; significant < 0.05).

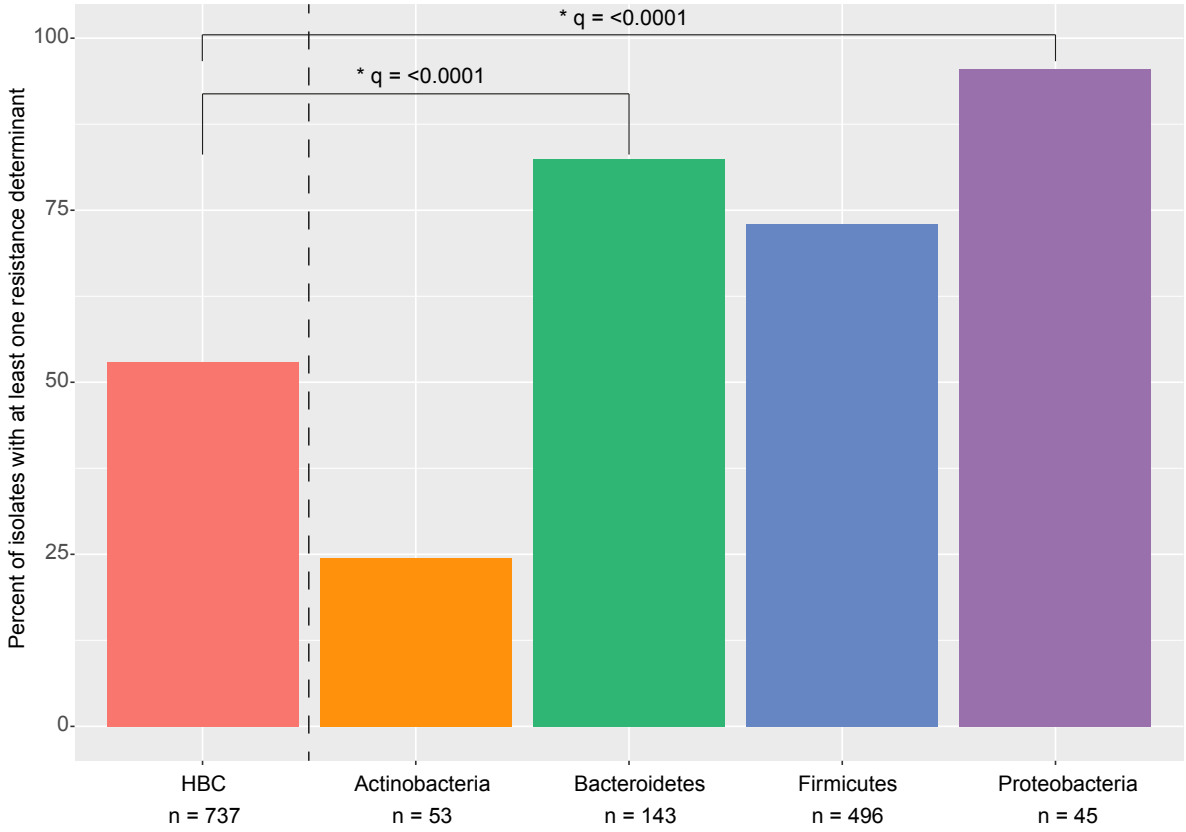

Figure 3.6. Proportions of isolates with at least one genetic antibiotic resistance determinant in each phyla, compared to the overall HBC. The numbers of genomes with at least one genetic antibiotic resistance determinant were counted for the complete HBC and for each of the four HBC phyla. HBC 52.9 %; Actinobacteria 24.5 %; Bacteroidetes 82.5 %; Firmicutes 43.5 %; Proteobacteria 95.6 %. Proteobacteria and Bacteroidetes had significantly more isolates with predicted resistance than expected based on the overall HBC collection. Statistical significance determined by Fisher exact tests, and corrected using the Benjamini, Hochberg, and Yekutieli method; q = significant < 0.05. n = total number of isolates in that group.

Moreover, Proteobacteria showed a bigger range and higher maximum number of antibiotic resistance determinants per isolate than the other three phyla (Fig. 3.7) and again, this was found to be statistically significant (q < 0.0001). Overall, the data from Figures 3.5 to 3.7 so far show that antibiotic resistance determinants are not distributed evenly between or within

phyla: more Bacteroidetes isolates are found to contain resistance determinants than expected, but the number of determinants per isolate was not significantly different to the HBC overall. The Proteobacteria phylum is significantly enriched for the number of isolates with resistance and the number of resistance determinants per isolate.

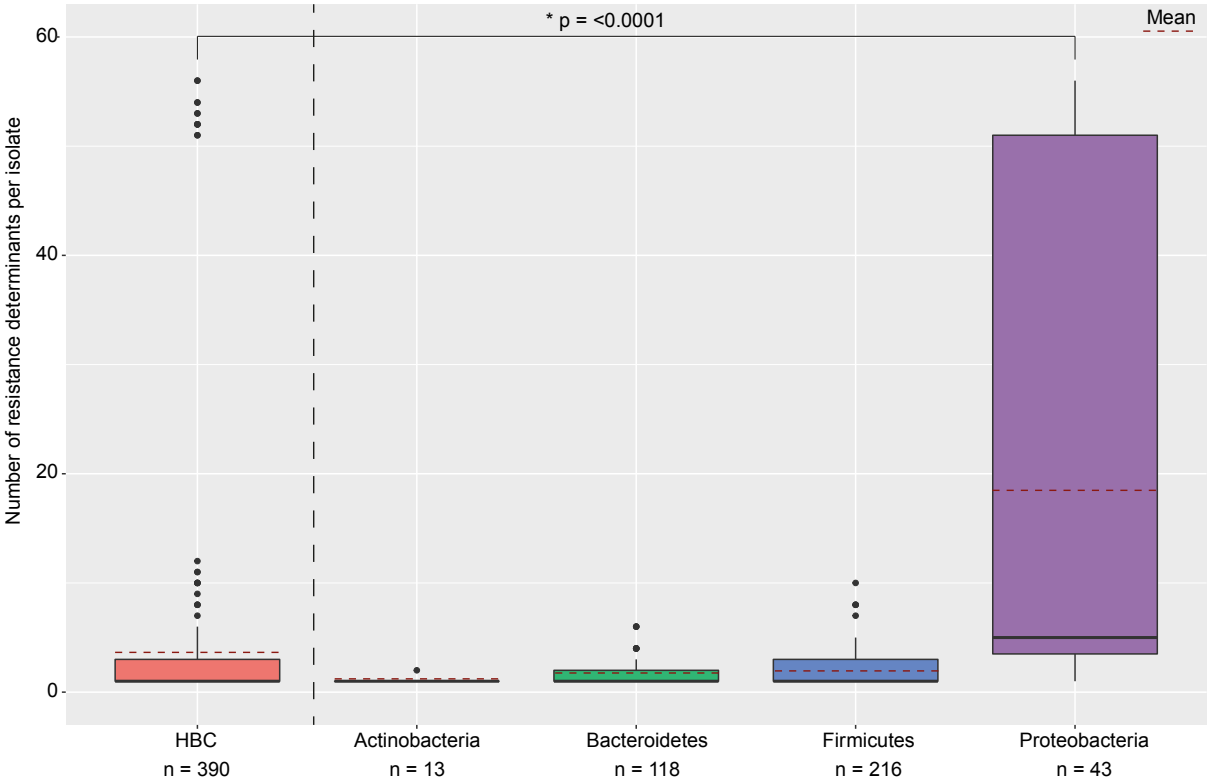

Figure 3.7. Interquartile range of number of antibiotic resistance determinants per isolate in each phyla, compared to the overall HBC. The total number of antibiotic resistance determinants was calculated for each isolate and the interquartile range in this number plotted for all 737 HBC genomes and for each of the four HBC phyla. Actinobacteria had the smallest range (0-2) and lowest total number of antibiotic resistance determinants per isolate of all the phyla. In contrast, Proteobacteria had the biggest range (0-56) and highest total number of antibiotic resistance determinants per isolate. Interquartile range was determined using the Tukey method; the black circles indicate outlier observations and the thick black line represents the median. The red dashed line represents the mean. The mean number of determinants per isolate was statistically higher in Proteobacteria than the HBC. Statistical significance determined by Mann-Whitney U test; p = significant < 0.05. n = number of isolates in that group with predicted resistance (i.e., excluding isolates without resistance).

Similarly, Proteobacteria display a bigger range in the number of specific antibiotic categories resistance is predicted to (i.e., excluding non-specific antibiotic resistances; Fig. 3.8): the Proteobacteria isolates are on average predicted to be resistant to four different antibiotics but could be resistant to up to 10 different antibiotics. Actinobacteria, Bacteroidetes and

Firmicutes are only predicted to be resistant to 1.0 or 1.5 antibiotics on average respectively. Again, these results were found to be statistically significant ($p < 0.0001$). This analysis excludes the non-specific antibiotic resistances. That the range in number of predicted resistances appears similar to the range in number of antibiotic resistance determinants per isolate suggests that non-specific antibiotic resistance is not a major factor for Actinobacteria, Bacteroidetes and Firmicutes. There is a bigger difference in the Proteobacteria, indicating that these isolates contain more non-specific resistances.

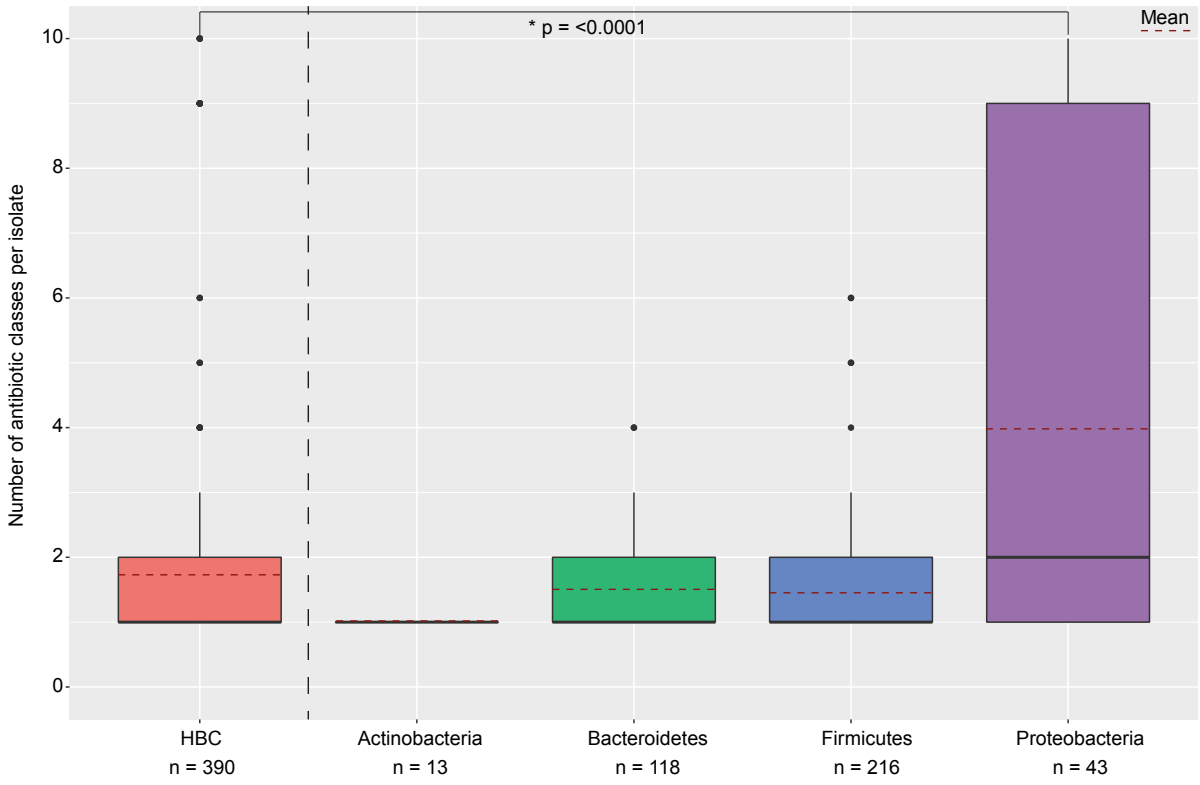

Figure 3.8. Interquartile range of number of antibiotic classes resistance is predicted to per isolate in each phylum, compared to the overall HBC. After grouping the determinants by the antibiotic class they are described as conferring resistance against by CARD, the total number of specific antibiotic classes a single isolate was predicted to harbor genetic resistances against was counted (i.e., excluding non-specific resistances), Again, Proteobacteria had the biggest range and highest maximum number of antibiotic classes per isolate. Interquartile range was determined using the Tukey method; the black circles indicate outlier observations and the thick black line represents the median. The red dashed line represents the mean. The mean number of different antibiotic classes per isolate was statistically higher in Proteobacteria than the HBC. Statistical significance determined by Mann-Whitney U test; p = significant < 0.05. n = number of isolates in that group with predicted resistance (i.e., excluding isolates without resistance).

Next, I considered whether certain phyla were enriched for particular antibiotic resistances. Fisher exact tests were performed to compare the proportion of isolates with resistance determinants for a particular antibiotic category within phyla compared to the proportion of all HBC isolates; p-values were adjusted for multiple testing using the Benjamini, Hochberg, and Yekutieli method (Table 3.2). Overall, every phylum except Actinobacteria is enriched for resistance to specific antibiotic classes. Bacteroidetes were enriched for beta-lactam resistance and tetracycline resistance (q values both < 0.001, significant < 0.05) and Firmicutes were enriched for diaminopyrimidine antibiotic resistances (q value 0.001). Proteobacteria were statistically enriched for determinants belonging to nine different classes of antibiotics, as well as non-specific antibiotic resistance determinants (q values all < 0.001). In several cases, resistance to antibiotic classes was observed significantly less often than expected – for example, beta-lactam resistance in Actinobacteria. This further indicates that genetic determinants of resistance are unevenly distributed between phyla. Overall, Proteobacteria are enriched for the most antibiotic classes and known resistance determinants, thus appearing to harbour more clinically relevant antibiotic resistance determinants than the other three key gut microbiota phyla.

Table 3.2. The proportion of isolates with resistance to particular antibiotic categories is compared between phyla and the overall HBC. Percentage represents proportion of predicted resistances for an antibiotic in the HBC or individual phyla. The individual phylum proportions were compared to the HBC proportion to determine the direction of statistical significance. The arrows denote the direction of significance; ↑ indicates that antibiotic resistance was observed more than expected and ↓ that antibiotic resistance was observed less than expected. No arrow means that any change in the proportion was not statistically significant. Q values significant < 0.05 (Fisher exact tests and adjusted using the Benjamini, Hochberg, and Yekutieli method).

		Phylum (number of genomes)											
		Actinobacteria (53)			Bacteroidetes (143)			Firmicutes (496)			Proteobacteria (45)		
Antibiotic	**HBC %**	**Phylum %**	**q-value***	**Direction**	**Phylum %**	**q-value***	**Direction**	**Phylum %**	**q-value***	**Direction**	**Phylum %**	**q-value***	**Direction**
Acridine Dye	0.3	0.0	1.000		0.0	1.000		0.4	1.000		0.0	1.000	
Aminocoumarin	1.6	0.0	1.000		0.0	1.000		0.0	< 0.001*	↓	26.7	< 0.001*	↑
Aminoglycoside	5.2	0.0	1.000		1.4	0.227		3.4	0.096		33.3	< 0.001*	↑
Beta-lactam	12.1	0.0	0.023*	↓	31.5	< 0.001*	↑	1.4	< 0.001*	↓	82.2	< 0.001*	↑
Chloramphenicol	0.5	0.0	1.000		0.0	1.000		0.8	1.000		0.0	1.000	
Diaminopyrimidine	5.2	1.9	1.000		0.0	0.007*	↓	7.3	0.001*	↑	2.2	1.000	
Elfamycin	1.1	0.0	1.000		0.0	1.000		0.0	0.002*	↓	17.8	< 0.001*	↑
Fluoroquinolone	1.6	0.0	1.000		0.0	1.000		0.0	< 0.001*	↓	26.7	< 0.001*	↑
Fosfomycin	3.5	0.0	1.000		0.0	0.066		0.4	< 0.001*	↓	53.3	< 0.001*	↑
Glycopeptide	1.5	0.0	1.000		0.0	1.000		2.2	0.227		0.0	1.000	
MLPS	9.4	0.0	0.087		11.9	1.000		14.1	1.000		6.7	1.000	
Mupirocin	0.1	1.9	0.767		0.0	1.000		0.0	1.000		0.0	1.000	
Nitrofuran	1.5	0.0	1.000		0.0	1.000		0.0	< 0.001*	↓	24.4	< 0.001*	↑
Nitroimidazole	1.6	0.0	1.000		0.0	1.000		0.0	< 0.001*	↓	26.7	< 0.001*	↑
Non-specific	11.8	5.7	1.000		0.0	< 0.001*	↓	8.7	0.004*	↓	91.1	< 0.001*	↑
Nucleoside	1.2	0.0	1.000		1.4	1.000		1.2	1.000		2.2	1.000	
Peptide	4.2	0.0	1.000		0.0	0.026*	↓	3.8	1.000		26.7	< 0.001*	↑
Sulphonamide	0.5	0.0	1.000		1.4	1.000		0.0	0.140		4.4	0.227	
Tetracycline	37.2	17.0	0.227		76.9	< 0.001*	↑	28.6	< 0.001*	↓	28.9	1.000	

3.2.4 Variation of predicted genomic resistance across different human commensal bacterial Families

Having identified significantly more antibiotic resistance overall in Proteobacteria so far, I wanted to see whether this was also true at the family level. Families with more than 5 isolates were selected to be analysed to allow a certain degree of robustness. Those 14 families were ranked by the proportion of isolates with predicted resistance (Fig. 3.9).

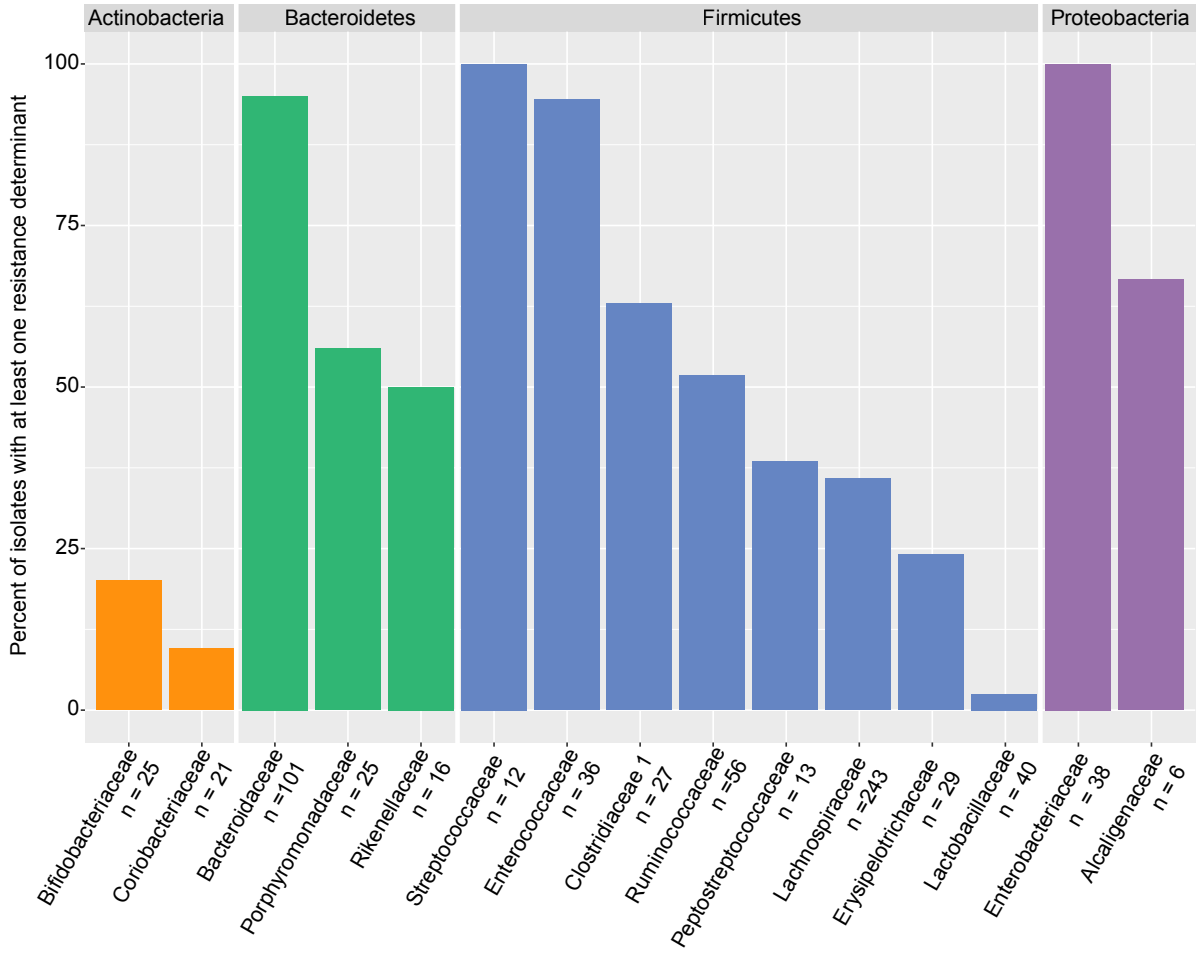

Figure 3.9. Proportions of isolates with at least one genetic antibiotic resistance determinant in taxonomic families. The numbers of genomes with at least one genetic antibiotic resistance determinant were counted for families in the HBC. The two families with the highest average number of resistance determinants per isolate or with more than five isolates are shown for each phylum.

This shows that as for phyla, families vary in their proportion of isolates with predicted antibiotic resistance. In particular, the highest rates of isolates with predicted antibiotic

resistance occur in Enterobacteriaceae, Streptococcaceae, Enterococcaceae, and Bacteroidaceae. These families are known to contain opportunistic pathogenic species. However, so is the family Peptostreptococacceae (namely, the diarrhoea-causing *Clostridioides difficile*), yet this family has a much lower proportion of isolates with predicted resistance. If we consider the number of genetic determinants per isolate between families (Fig. 3.10), even though Streptococcaceae and Bacteroidaceae have high proportions of isolates with predicted resistance, they only harbour relatively few resistance determinants per isolate. In comparison, Enterococcaceae and Enterobacteriaceae both have significantly more resistance determinants per isolate on average compared to Bacteroidaceae and Streptococcaceae. However, Enterobacteriaceae did not have significantly more resistance determinants on average than Enterococcaceae. Together so far, this data suggests that gut microbiota taxa known to contain isolates of species that can act as opportunistic pathogens are enriched for genetic determinants of antibiotic resistance.

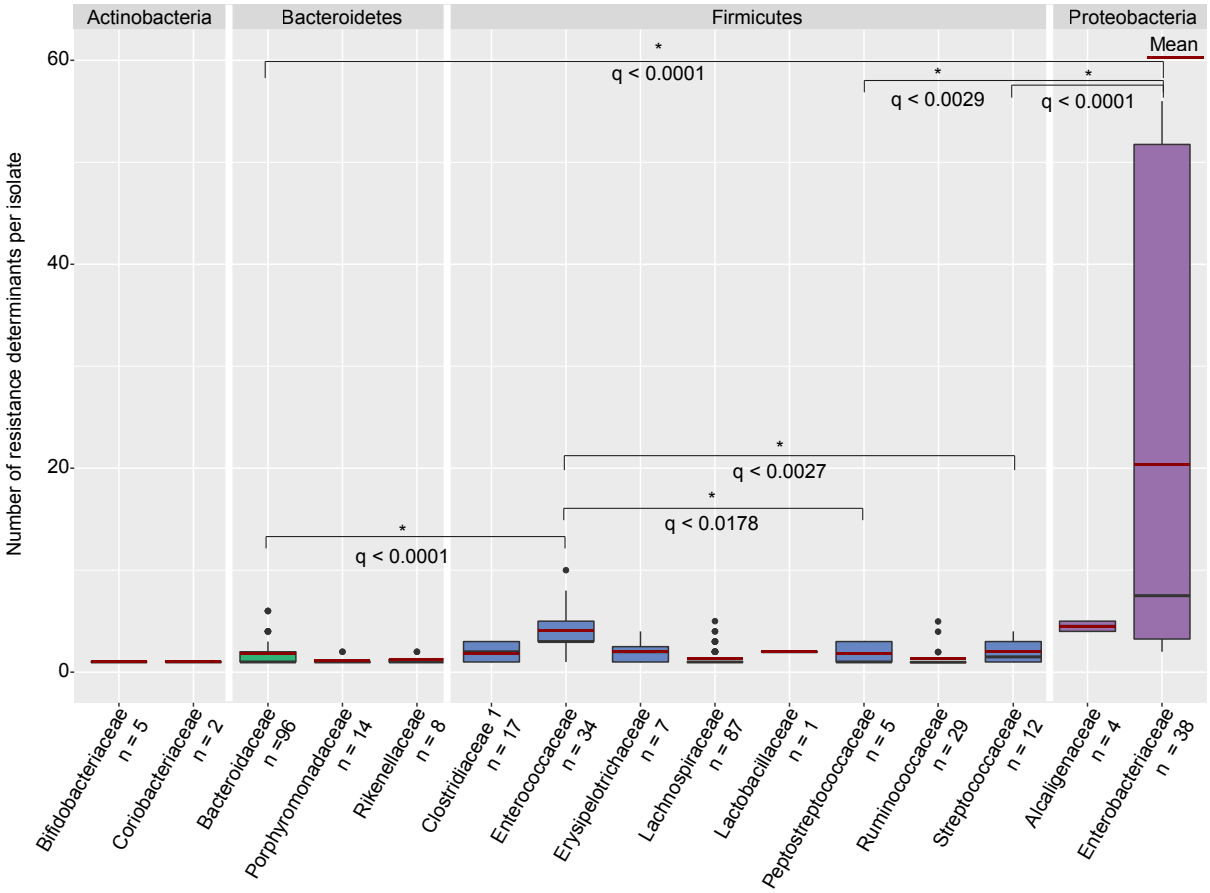

Figure 3.10. Interquartile range of number of antibiotic resistance determinants per isolate in commensal gut bacterial families. The total number of antibiotic resistance determinants was calculated for each isolate and the distribution in this number plotted at the family level. This included non-specific antibiotic resistance determinants. Interquartile range was determined using the Tukey method; the points show outlier observations and the thick black line represents the median. The red line represents the mean. n = number of isolates in that group with predicted resistance. Only families with more than five isolates were included. Statistical significance was determined between families known to contain pathogenic bacteria (Bacteroidaceae, Enterococcaceae, Peptostreptococacceae, Streptococcaceae and Enterobacteriaceae) by Kruskal-Wallis tests and corrected using the Benjamini, Hochberg, and Yekutieli method; q = significant < 0.05. Only significant results are shown.

3.2.5 Distribution of predicted genomic resistance between known and novel isolates

So far, I have identified that families containing species that can act as opportunistic pathogens are enriched for genetic determinants of antibiotic resistance. Having such a diverse and novel collection of gut bacteria has also allowed me to investigate whether clinically relevant antibiotic resistance determinants are harboured not just by bacteria belonging to previously published or described (“known”) taxa, but those considered to be novel and thus uncharacterised. As we do not have novel isolates for many of the families in

the HBC, I performed this analysis at the phylum level. Fewer novel genomes overall harbour clinically relevant antibiotic resistances (Fig. 3.11); this is also true for the Bacteroidetes phylum. There were no novel Actinobacteria in this dataset so the rates in this phylum could not be compared. In Proteobacteria and Firmicutes (which has the most novel genomes at 253), it is the novel genomes that harbour more clinically relevant antibiotic resistances. However, the only statistically significant results were that the HBC overall and Bacteroides phyla had more known isolates with resistance than novel isolates; the observations in the Firmicutes were not significantly different. Nonetheless, these results indicate that uncharacterised bacteria contribute to the intestinal reservoir of known antibiotic resistance.

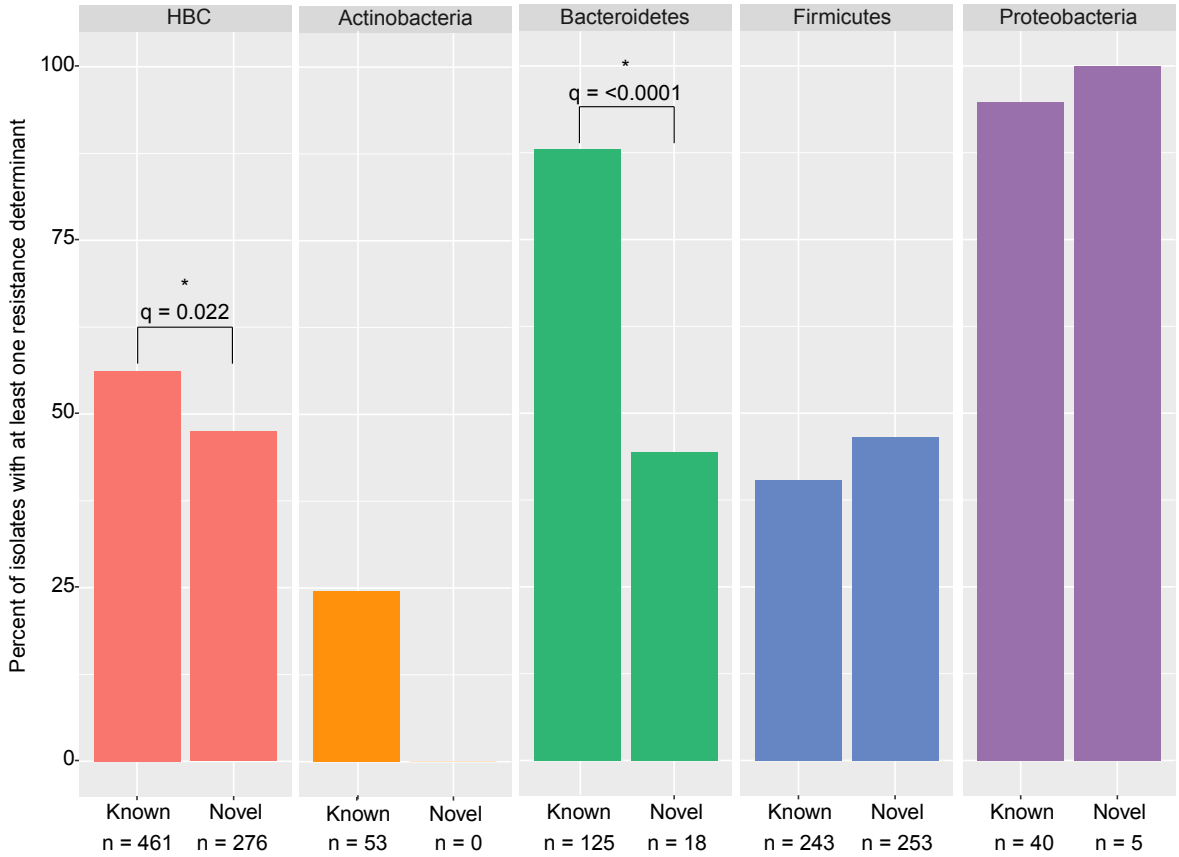

Figure 3.11. Proportions of resistant isolates in known versus novel isolates. The numbers of isolates that were novel were counted for all 737 HBC genomes and for each of the four HBC phyla. For each sub-group (known or novel), isolates with at least one genetic antibiotic resistance determinant were counted as before. 47.4 % of the novel genomes in the HBC were predicted to have at least one genetic antibiotic resistance determinant overall. There were no novel Actinobacteria genomes; the majority (91.7 %) of novel genomes belong to Firmicutes, with 46.6 % of novel Firmicutes predicted to have at least one genetic antibiotic resistance determinant overall. Overall, the HBC had significantly more known isolates with predicted resistance than novel isolates; within phyla only Bacteroides showed a significant difference. Statistical significance determined by Fisher exact tests, and corrected using the Benjamini, Hochberg, and Yekutieli method; q = significant < 0.05. n = total number of isolates in that group.

Overall, the known genomes of the HBC harboured resistances to all 19 categories of antibiotics, whereas the novel genomes only harboured resistances to 12 categories of antibiotics (Figure 3.12, showing the proportion of predicted resistances). Known and novel Bacteroidetes genomes harboured resistances to the same six antibiotics, although there was proportionately more diversity in the resistances of novel Bacteroidetes. Similarly, novel Firmicutes harboured resistances to the same 11 categories of antibiotic as known Firmicutes, though there were more tetracycline resistances in the novel Firmicutes compared to known Firmicutes. The proportion of resistances in known and novel Bacteroidetes and Firmicutes appears broadly similar, largely dominated by tetracycline resistance. In contrast, the Actinobacteria and Proteobacteria have different patterns in predicted resistances: the known Actinobacteria have resistances to all 19 classes of antibiotics and known Proteobacteria to 14 classes of antibiotics, whereas the novel Proteobacteria only have eight classes of antibiotic resistance predicted. In addition, the novel Proteobacteria have different classes of antibiotic resistance predicted (i.e., acridine dye, diaminopyrimidine, glycopeptide) than the known Proteobacteria. Together, these observations make it clear that although Proteobacteria is enriched for antibiotic resistance, non-Proteobacteria and uncharacterised gut microbiota harbour diverse antibiotic resistances, underlining the importance of understanding this extensive reservoir and its clinical relevance.

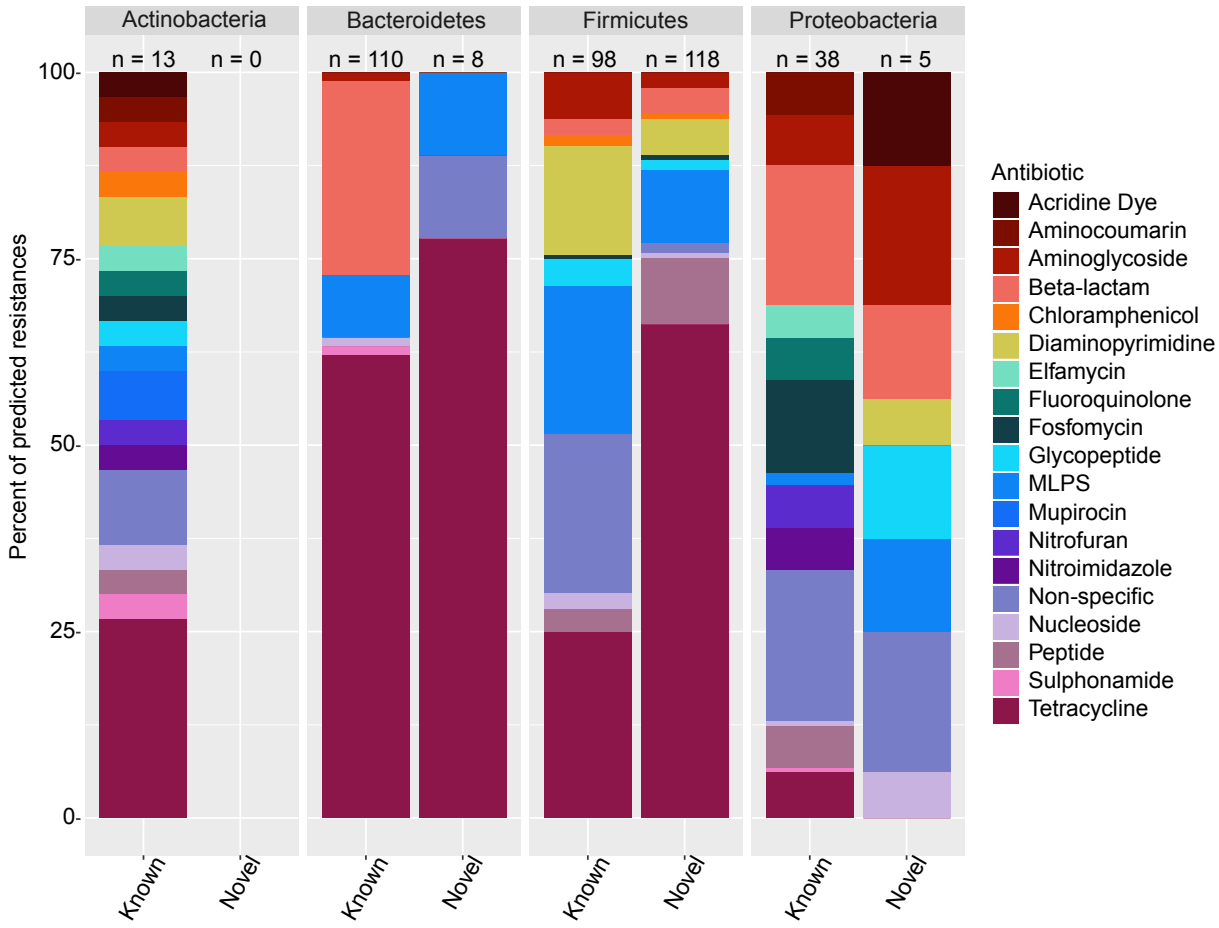

Figure 3.12. Proportions of antibiotic categories that resistances are predicted against in HBC genomes. The number of predicted resistance phenotypes in known and novel genomes per antibiotic class was calculated for the HBC and each phylum. The majority of predicted resistances in those novel Firmicutes were against tetracycline antibiotics. r = total number of predicted resistance phenotypes for that group; g = genomes with predicted resistance in that group.

3.2.6 Comparison of predicted genomic resistance in commensal versus pathogenic isolates

I next compared the presence of antibiotic resistance in commensals of the HBC to pathogens. To do this, I searched PATRIC[194] 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 bacterial infections, and there are commensal isolates of these species in the HBC for comparisons. 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

contemporaneous 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 *Enterococcus faecium* isolates with official isolation dates post-2010, but looking at other meta-data 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. For the 1725 genomes, I predicted the presence of genetic antibiotic resistance determinants in CARD using ARIBA as for the HBC.

Importantly, 97 % of these pathogenic genomes had at least one predicted resistance. However, there is more variation in the number of resistance determinants per isolate between species (Fig. 3.13) than was seen for the commensal isolates in section 3.2.4. This shows that although resistance is predicted in the majority of pathogenic isolates, different pathogenic species of bacteria have different antibiotic resistance potential. Pathogenic *E. coli* has the highest maximum, median and mean number of determinants per genome; this is similar to the observations in commensal HBC Proteobacteria.

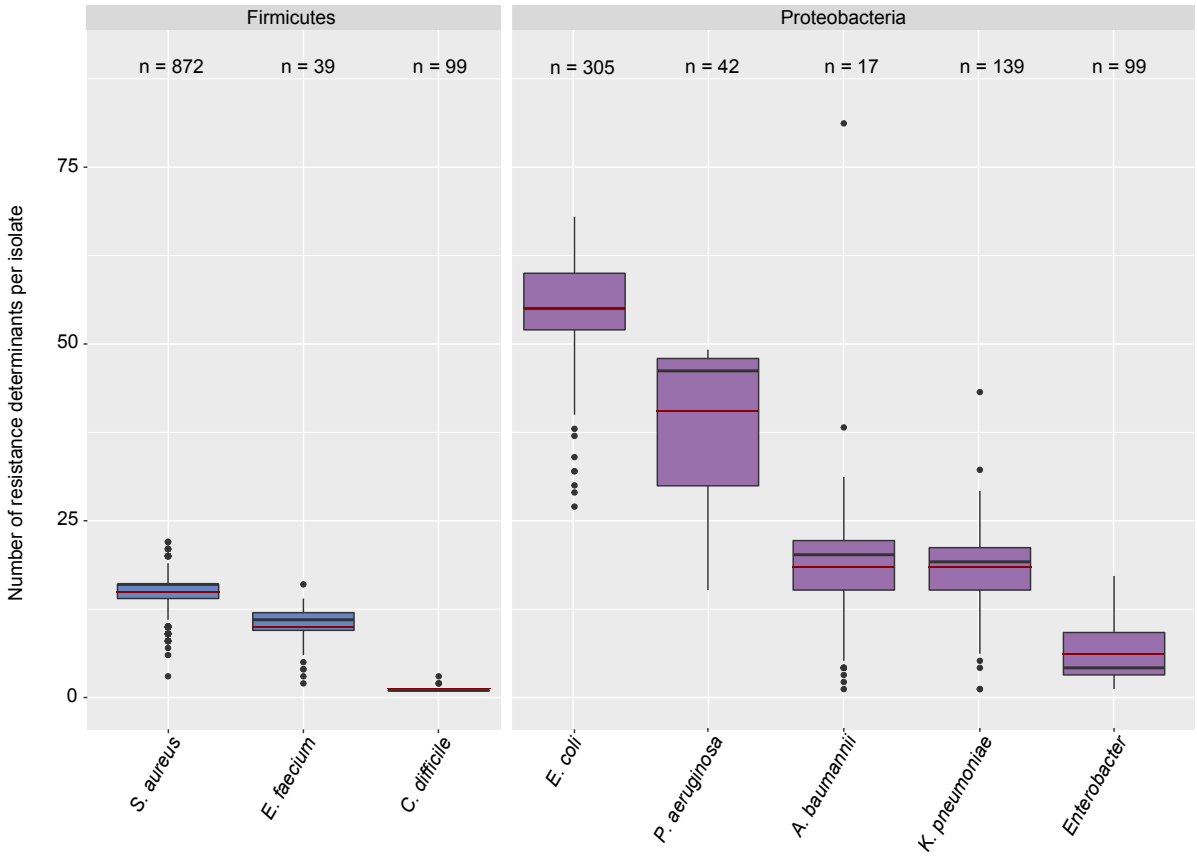

Figure 3.13. Interquartile range of number of antibiotic resistance determinants per isolate in pathogenic bacterial species. The total number of antibiotic resistance determinants was calculated for each pathogenic genome studied and the distribution in this number plotted at the species level. This included non-specific antibiotic resistance determinants. n = number of isolates in that group with predicted resistance. Interquartile range was determined using the Tukey method; the points show outlier observations and the thick black line represents the median. The red line represents the mean.

In addition, different pathogenic species have different profiles of predicted resistances (Fig. 3.14). Thus, not all of these pathogen groups are equal in their predicted propensity to harbour antibiotic resistances. The Proteobacteria species and *S. aureus* appear broadly similar, with eight or more different resistances predicted. In contrast, *C. difficile* and *E. faecium* appear less similar, with only three and seven categories of antibiotic resistances predicted respectively.

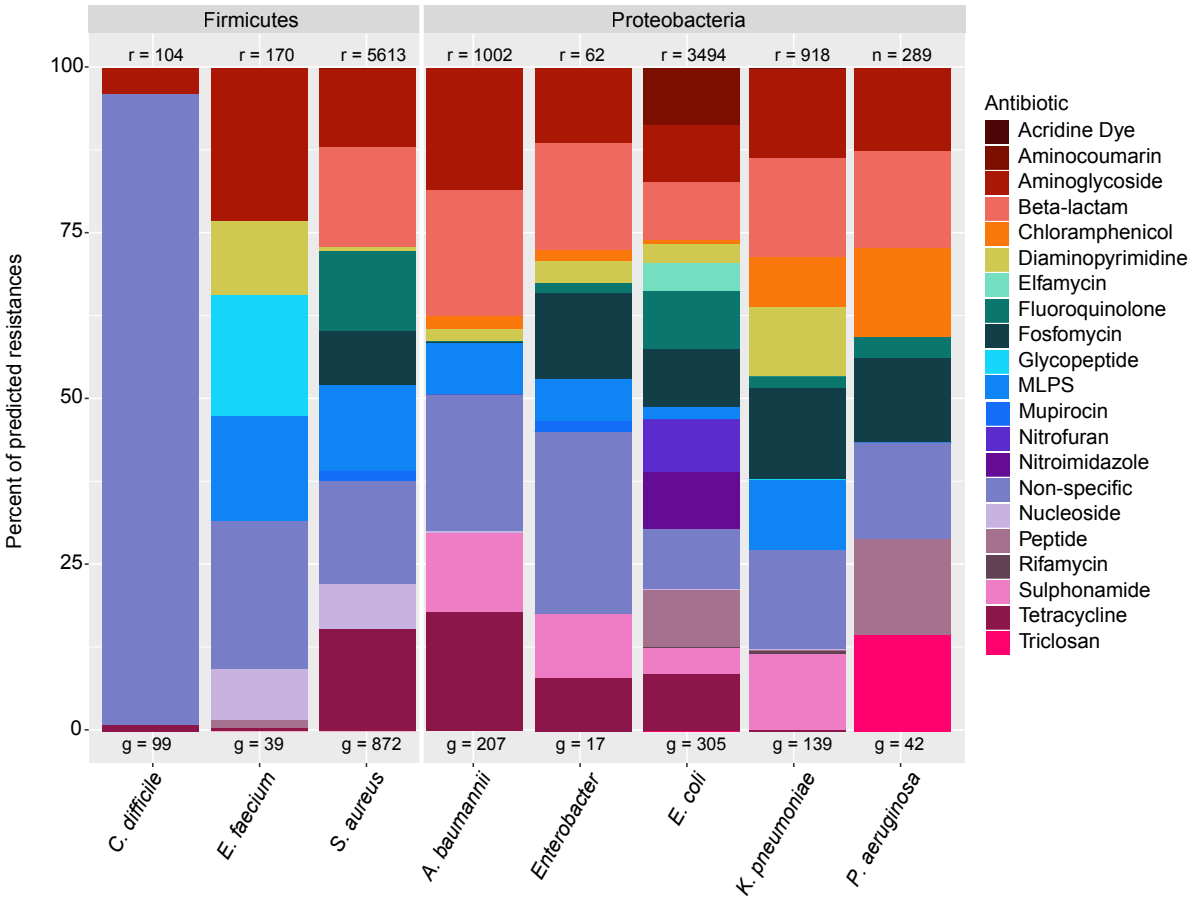

Figure 3.14. Proportions of antibiotic categories that resistances are predicted against in pathogenic bacterial genomes. The percentage of predicted resistance phenotypes in pathogenic genomes per antibiotic class was calculated for each species. r = total number of predicted resistance phenotypes for that group; g = genomes with predicted resistance in that group.

Having assessed the frequency and proportions of genomic resistance in certain pathogenic species, I wanted to directly compare between pathogenic and commensal (from the HBC) genomes of related species. I have performed this analysis at the family level for Enterococcaceae (includes pathogenic *E. faecium*), Peptostreptococacceae (includes pathogenic *C. difficile*), Staphylococcaceae (includes pathogenic *S. aureus*), and Enterobacteriaceae (includes pathogenic *Enterobacter*, *E. coli* and *K. pneumoniae*) due to limitations in numbers of HBC genomes for certain species. Comparing the number of determinants per genome directly between commensal HBC and pathogenic isolates of the

same families, we can see that the pathogenic isolates have significantly more determinants per genome on average, except for Peptostreptococcaceae (Fig. 3.15).

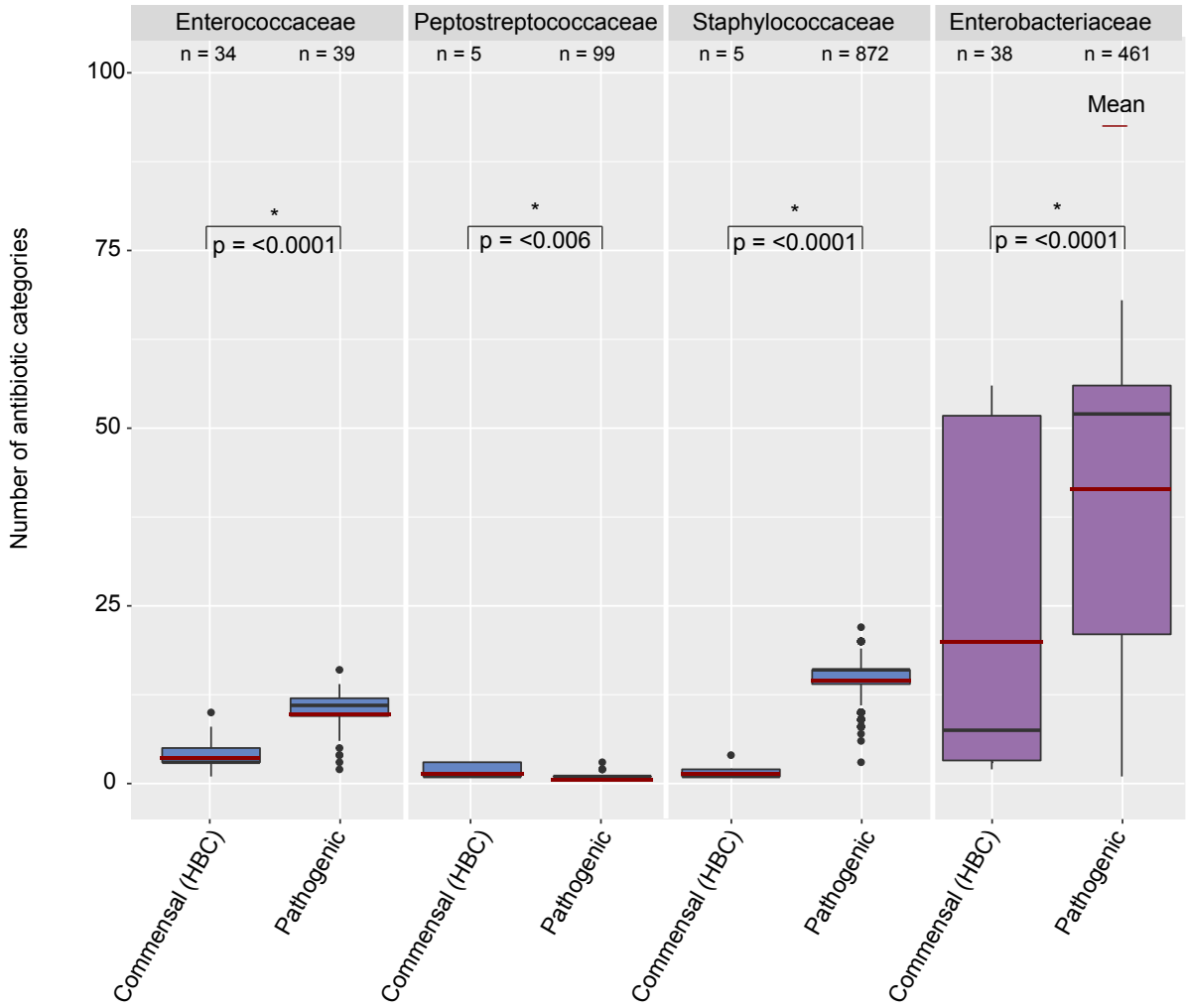

Figure 3.15. Interquartile range of number of antibiotic resistance determinants per isolate in commensal versus pathogenic isolates. The total number of antibiotic resistance determinants was calculated for each genome studied and the distribution in this number plotted at the family level, where families were represented by more than one isolate in both datasets. n = number of isolates in that group with predicted resistance. Interquartile range was determined using the Tukey method; the black circles show outlier observations and the thick black line represents the median. The red line represents the mean. Statistical significance determined by Mann-Whitney U test; p = significant < 0.05.

However, we can also see that Enterococcaceae and Enterobacteriaceae each have broadly similar resistance profiles in both commensal HBC and pathogenic isolates of these families (Fig. 3.16). In contrast, pathogenic Staphylococacceae have more categories of antibiotic resistances predicted than commensal HBC isolates. Peptostreptococacceae demonstrate a

different pattern again: the pathogenic isolates are dominated by non-specific resistances, whereas the commensal isolates have more tetracycline and MLPS resistances predicted. This data shows that although pathogenic isolates of Enterococcaceae and Enterobacteriaceae carry more resistance determinants, overall commensal and pathogenic isolates in these families are predicted to be resistant to similar antibiotics. In contrast, commensal isolates of Peptostreptococacceae and Staphylococcaceae are predicted to be resistant to different antibiotics than pathogenic isolates.

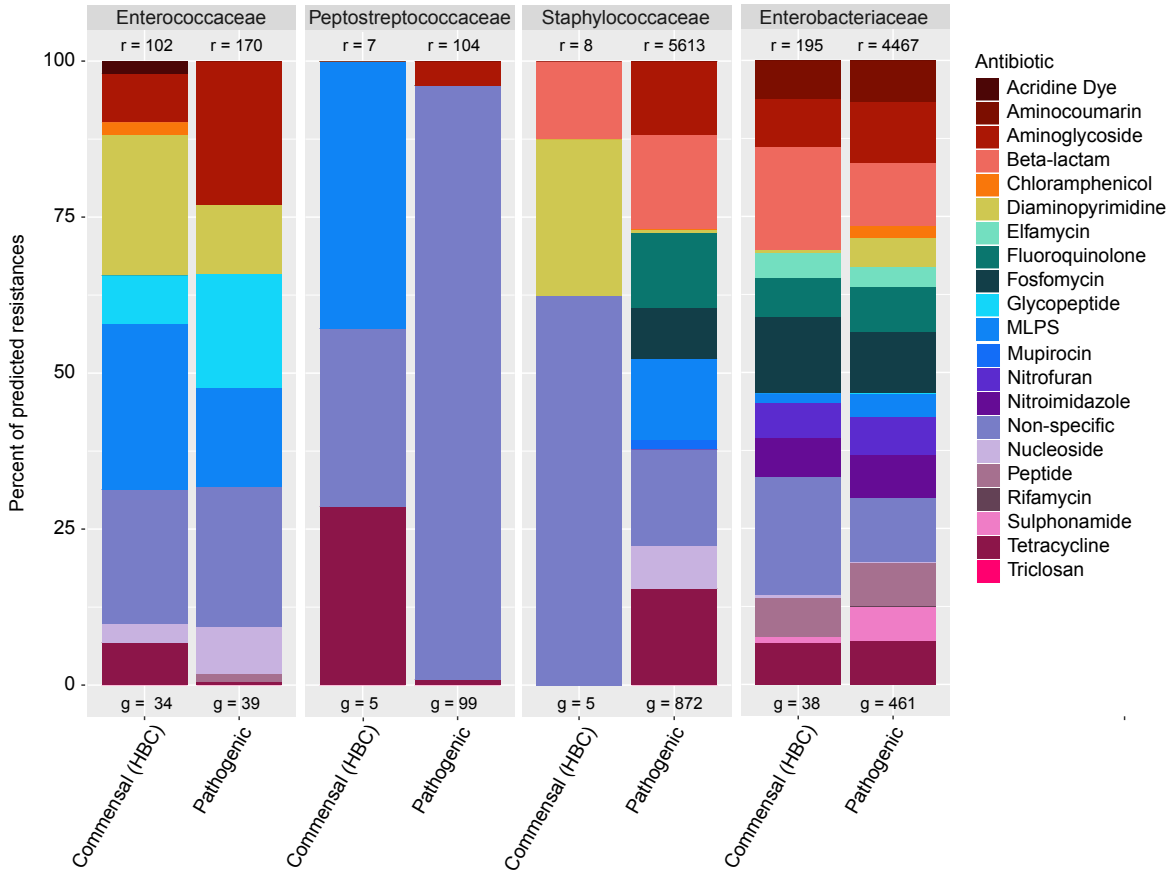

Figure 3.16. Proportions of antibiotic categories that resistances are predicted against in commensal HBC versus pathogenic bacterial genomes. The number of predicted resistance phenotypes in genomes per antibiotic class was calculated for each family. r = total number of predicted resistance phenotypes for that group; g = genomes with predicted resistance in that group.

3.3 Discussion

In this chapter, I aimed to determine the presence of known antibiotic resistance determinants in pure isolates of commensal gut bacteria and see if the whole gut microbiome may contribute to the problem of antibiotic resistance or only certain members. I therefore screened a collection of 737 phylogenetically diverse human gut bacterial genomes for known antibiotic resistance determinants described by CARD. This identified 178 unique antibiotic resistance determinants across the HBC, predicted to confer resistance against 19 antibiotic categories (including a category for non-specific antibiotic resistance). This included all but one of the antibiotics on the WHO Essential Medicines List, oxazolidinone. As in previous studies[140,142], tetracycline resistance genes were some of the most prevalent across the diversity of the HBC.

There were nine other antibiotics described in CARD to which there were not any predicted specific resistances: rifamycin, polyamine, triclosan, isoniazid, para-aminosalicylic acid, antibacterial free fatty acids, bicyclomycin and pyrazinamide (listed in order of the number of resistance determinants against these antibiotics in CARD). Rifamycin has broad-spectrum activity against Gram-negatives and Gram-positives, but is mainly used to treat TB infections[223]. There are no *Mycobacterium* isolates in the HBC, which we would expect given the criteria for the human donors for the HBC study to be healthy and without any bacterial infections. Rifamycin resistance develops by mutations in the RNA polymerase gene, and if the donors had not received rifampicin treatment recently then this may explain the absence of rifamycin-conferring resistance mutations in genomes of gut bacteria isolated from those people. Isoniazid[224], para-aminosalicylic acid and pyrazinamide are also used to treat TB and so a similar explanation may apply. However, there were non-specific determinants present in these genomes that included rifamycin as one of the antibiotics they are described as

conferring resistance to (Appendix II). Similarly, triclosan and oxazolidinone were included in some non-specific determinant descriptions. Bicyclomycin is an old antibiotic with weak activity that is being revived for combinatorial therapy of drug resistant infections[225]. Polyamine is a compound that can increase the susceptibility of bacteria to other antibiotics[226], such as beta-lactams, that is also being developed as an option combinatorial therapy[227]. Antibacterial free fatty acids are another experimental alternative to antibiotics[228]. Since these antibiotics are not currently commonly used as treatment of bacterial infections, resistance is less likely to have had the chance to develop or spread as often as resistance to other, more commonly used antibiotics. This could explain the absence of resistance determinants to these three antibiotics in the HBC.

Overall, resistances were enriched among the Proteobacteria, and particularly the Enterobacteriaceae, in terms of diversity of resistance determinants, abundance of resistance determinants per isolate, and the number of isolates with predicted resistance. Isolates of Enterobacteriaceae harboured between 1 and 31 non-specific resistance determinants; this means the Proteobacteria could be resistant to more antibiotics than assumed when just considering the specific antibiotic resistance determinants. Since the HBC isolates were isolated from the guts of healthy adults who had not taken antibiotics in at least six months, and the microbiome is a diverse population of microorganisms, it is unlikely that the Enterobacteriaceae isolates have been more exposed to antibiotics than non-Proteobacteria isolates. As many of the most important bacterial pathogens are Gram-negatives – often Enterobacteriaceae or other Proteobacteria – some of the most commonly used antibiotics are those that target Gram-negatives. Therefore, it is possible that when taking antibiotics the commensal Enterobacteriaceae are theoretically more likely to be impacted and thus under more selective pressure to develop antibiotic resistance. If this were maintained long-term

(beyond the period of antibiotic treatment), this may explain the observations in this chapter. This is plausible, however there is often a fitness cost associated with antibiotic resistance[229], implying that any acquired antibiotic resistance genes or mutations are not likely to be maintained in the absence of antibiotic selective pressure.

Alternatively, it is possible that the HBC Enterobacteriaceae acquired resistance genes from bacteria in the environment or diet, in line with the "One Health" concept discussed in the Introduction to this thesis, and that this has happened more commonly in this taxon. However, horizontal gene transfer is not necessarily more common in Proteobacteria than other phyla: a recent study found that phylogeny did not have a significant impact on genome fluidity[230]. Conversely, and more likely, these results may reflect how the databases are predominantly created from information generated by studying bacterial pathogens; throughout this chapter it is the families of bacteria known to contain pathogenic species that have the most antibiotic resistance predicted. The databases would therefore be biased towards predicting antibiotic resistance genes or mutations in bacterial genomes more similar to those studied for the creation of said databases. This bias could explain the enrichment of resistance in Enterobacteriaceae over more distantly related bacteria, such as the Firmicutes.

Despite the apparent enrichment in Proteobacteria, there is still a diverse range of antibiotic resistance predicted in large proportions of the other phyla in the HBC. For example, Firmicutes were enriched for diaminopyrimidine (trimethoprim) resistance; *dfr* trimethoprim resistance genes have been described in both Gram-negative and Gram-positive bacteria, but are thought to be intrinsic to enterococci[231]. Moreover, Bacteroidetes were enriched for beta-lactam resistance; beta-lactamases are often endogenously produced by these bacteria[232]. Finally, tetracycline resistance was enriched in Bacteroidetes, and was the most common type of resistance in the HBC. Tetracycline resistance has been reported as increasingly common in

Bacteroidetes[233] and the gut microbiota generally[123], and serves to highlight how antibiotic resistance can become widespread in commensal bacteria. Whilst tetracyclines are now infrequently used in human medicine for this reason, there is a possibility of cross resistance with the last resort glycylcycline antibiotic tigecycline, and so the presence of tetracycline resistance determinants remains relevant. There were also instances of resistance to antibiotic classes being observed significantly less often than expected, such as to beta-lactams in the Actinobacteria isolates; this would suggest either the isolates of those phyla are all susceptible to that antibiotic, or that there are alternative resistance mechanisms not identified using the methods applied here.

Many of the predicted resistances in novel HBC genomes, most of which belong to the phylum Firmicutes, were to tetracycline. Other resistances were also predicted such as MLPS, beta-lactams and diaminopyrimidine (trimethoprim). If known resistance genes were not identified in novel isolates, then the fact there are still gut bacteria to be discovered[216] does not necessarily have serious consequences for the problem of antibiotic resistance. However, almost half of the novel HBC isolates did contain antibiotic resistance genes or mutations. Since there are uncharacterised gut bacteria, this emphasises that we do not yet fully understand the gut microbiota and the extent of their contribution to antibiotic resistance. Overall, significantly more known HBC genomes contained resistance determinants than novel HBC genomes; the same statement is true for the Bacteroidetes phylum specifically. There was no significant difference between the numbers of known or novel Firmicutes or Proteobacteria with resistance determinants. For the Proteobacteria, it is possible that this is due to sample bias, as there were only five novel Proteobacteria isolates, whereas in Firmicutes it is almost a 50:50 division of novel and known genomes. This suggests that known bacteria, which are more likely to be more closely related to known pathogenic bacteria, are

more likely to contain antibiotic resistance, and that more distantly related (novel and uncharacterised) bacteria are less likely to contain antibiotic resistance. However, it is possible that again database bias is impacting these results, and that known bacteria are more likely to be predicted to contain antibiotic resistance than more distantly related bacteria. Overall though, this data emphasises that commensal gut microbiota do harbour clinically relevant antibiotic resistances and that this reservoir, much of which remains uncharacterised, is more extensive than currently realised.

Antibiotic resistances were commonly predicted throughout the HBC and there was variation observed even between closely related isolates; these statements are known to be true for pathogens as well as commensals. The commensal families most enriched for genetic determinants of antibiotic resistance are those known to contain opportunistic pathogenic species or isolates, such as Enterobacteriaceae and Enterococcaceae. In the last part of this chapter, I studied the differences in antibiotic resistance genotypes between commensal and pathogenic bacteria. I found that resistance determinants were observed more frequently in pathogenic isolates, and again especially in Proteobacteria.

As mentioned, antibiotics are often targeted towards these bacteria since many pathogenic species are Gram-negative Proteobacteria (Fig. 1.3); it is possible then that these bacteria are more exposed to antibiotics and thus more likely to develop antibiotic resistance. This may be partially true, but antibiotics can have “off-target” effects on other bacteria; other antibiotics are broad-spectrum and thus designed to target several types of bacteria, including Gram-positives. This means that other members of the gut microbiome will also be impacted by antibiotic use and thus also under selective pressure to develop resistance. Potentially, Proteobacteria might be more genetically capable of developing antibiotic resistance: Gammaproteobacteria (which includes Enterobacteriaceae) are known to have super-

integrons as an integral component in their genomes[234], allowing for multiple-drug resistance to develop rapidly. However, this concept of increased HGT in particular taxa was discussed earlier in this section and is unlikely to be the case. Alternatively, the enrichment of antibiotic resistance determinants in Proteobacteria may be due to efflux pumps being more common in Gram-negative bacteria than in Gram-positives[235]. Here, the majority of resistance determinants in the Proteobacteria were non-specific, of which efflux mechanisms dominated. This confirms that these non-specific mechanisms are a major contributor to the antibiotic resistance potential in Proteobacteria and may explain why pathogenic Proteobacteria are so prone to cause multi-drug resistant infections.

However, the average number of determinants per genome was lower in *Enterobacter* spp. than in *E. faecium* and *S. aureus*. Since *Enterobacter* spp. are Proteobacteria, this is perhaps unexpected based on my earlier findings, but could be explained by the fact that these bacteria have only relatively recently been described as an emerging multi-drug resistant threat[236]. In addition, pathogenic *E. faecium* and *C. difficile* genomes had resistances predicted to fewer antibiotic categories than pathogenic *S. aureus* or the Proteobacteria species. Pathogenic Gram-negative Proteobacteria are known for having mobile genetic elements that can contain several resistance genes at once that can confer resistance to different antibiotics[237]. In addition, pathogenic *S. aureus* is well-known for being resistant to multiple antibiotics[238], despite being Gram-positive. This may explain the difference in predicted resistance profiles observed.

When I compared the pathogenic versus commensal genomes, pathogenic isolates were generally found to have more genetic resistance determinants per genome on average than commensal isolates, particularly Enterobacteriaceae. The exception was for Peptostreptococcaceae, although it is possible that this is due to the number of genomes

studied (five commensals versus 99 pathogens) and studying additional genomes of commensal Peptostreptococcaceae will determine whether this pattern continues to be seen. In general for the comparisons between commensals and pathogens, the sample bias means these results should be treated with caution.

Finally, I found that for Peptostreptococacceae and Staphylococacceae, commensal genomes had different profiles of predicted antibiotic resistances to pathogenic isolates, whereas Enterobacteriaceae and Enterococcaceae shared similar predicted resistance profiles between commensal and pathogenic isolates. It is possible that this could be explained by more frequent horizontal sharing of resistance determinants within Enterobacteriaceae or Enterococcaceae than other bacterial families, though as discussed above phylogeny has not been observed to impact this process[230]. However, that study was performed at the species level, and so in the future it would be interesting to more specifically estimate and compare the frequency of horizontal gene transfer at other taxonomic levels, such as within and between bacterial families. This will also help to understand the spread of antibiotic resistance better. Yet again, it is important to consider the potential bias of the database of antibiotic resistance genes and mutations. This bias could also explain the observed higher proportions of pathogenic species with predicted resistance (and the high numbers of resistance determinants per genome) compared to commensal HBC isolates: the pathogenic isolates will be more similar to the isolates studied and used to create the database.

It is important to acknowledge that this study has only used one collection of commensal gut bacteria and one database of antibiotic resistance determinants; repeating these analyses with additional genomes and alternative databases will help confirm these findings. However, since the CARD database is a regularly updated and extensive collection of antibiotic resistance determinants, and the HBC is a recent and diverse collection of gut bacteria, these

new findings add to current knowledge regarding the gut microbiome as a reservoir of antibiotic resistance. The most important message from this chapter is that antibiotic resistance genotypes are common in commensals and often share similar predicted resistance profiles to related pathogens. In addition, many of the resistance genotypes are to antibiotics on the WHO list of essential medicines[21], emphasising the potential clinical relevance of these observations.

As the findings in this chapter are based purely on predicted genotypes using known antibiotic resistance determinants, this does not necessarily preclude that the isolates studied here are not resistant to other antibiotics – or even that they are resistant to the ones predicted. As discussed, there may be a database bias that makes it more likely to predict antibiotic resistance in bacteria more closely related to pathogenic species. Only phenotypic testing of antibiotic sensitivity of HBC isolates will confirm whether or not the observations in this chapter are accurate predictions of phenotypic antibiotic resistance. I will investigate this in the next chapter.