Chapter 4: Determination of phenotypic antibiotic resistance in commensal gut bacteria and the accuracy of genomic predictions

4.1 Introduction

4.1.1. Overview

In the previous chapter, I identified the presence of a range of clinically relevant genetic antibiotic resistance genes and mutations in the 737 genomes of the HBC. This analysis identified a significant enrichment of ARGs in the Proteobacteria members of the HBC. In this chapter I investigate whether Proteobacteria are phenotypically enriched for antibiotic resistance using the HBC culture collection, or whether CARD is biased towards identifying resistance genotypes in this phylum. In addition, I assess how accurate these predictions of antibiotic resistance in commensal gut bacteria are by determining phenotypic susceptibility and resistance to a range of clinically relevant antibiotics in a subset of the HBC.

The gold-standard method for determining isolate-specific and phenotypic antibiotic susceptibility is culture-based antibiotic susceptibility testing (AST). Developed mainly for clinical isolates of pathogenic bacteria, AST involves culturing the bacterium of interest in the presence of an antibiotic and observing its ability to grow. One common method is measuring the size of a zone of inhibition: this features a paper disk containing a single concentration of the antibiotic that is placed on an agar plate that has been inoculated all over with the isolate of interest – after 24 hours incubation this produces a bacterial lawn. The sensitivity of the isolate to that antibiotic determines how close to the disk it can grow: the more sensitive the isolate, the less close to the disk it will be able to grow. The diameter of the zone where no growth occurs (the zone of inhibition) is measured. The antibiotic disks are 0.5cm in diameter,

allowing several disks to be placed on a single inoculated plate and many antibiotic and isolate combinations to be tested using few resources[239]. Another method includes measuring the minimum inhibitory concentration (MIC). This involves exposing a bacterial isolate to a stepwise increasing range of antibiotic concentrations. This can be performed using broth microdilution (a series of prepared liquid culture mediums each with a different antibiotic concentration is inoculated with the isolate of interest) or using antibiotic gradient strips. These are small rectangular paper strips that contain a gradient of an antibiotic and are placed on top of an agar plate that has been inoculated to produce a bacterial lawn. The antibiotic diffuses into the agar – the more concentrated end will diffuse further – and then after 24 hours incubation the antibiotic concentration at which the isolate is no longer able to grow alongside the strip (the MIC), is measured.

4.1.2. Defining isolates as antibiotic-susceptible or -resistant

The zone of inhibition or MIC is compared to guidelines provided by the European Committee on Antimicrobial Susceptibility Testing (EUCAST) or the Clinical and Laboratory Standards Institute (CLSI); international agencies who study and determine zone of inhibition breakpoints that categorise a certain isolate as resistant, susceptible or having intermediate sensitivity. This depends on having culturable, purified isolates of the bacteria of interest – which, until recently has been difficult to achieve for the majority of gut bacteria. Therefore, these breakpoint guidelines are only available for a small number of pathogenic bacteria species.

As the costs of DNA sequencing decrease, whole genome sequencing antibiotic susceptibility testing (WGS-AST) is being investigated as an alternative method to culture-based AST to

determine antibiotic sensitivity. For example, the genome of a disease-causing bacterium isolated from a patient is sequenced and then searched for known antibiotic resistance determinants; any found are assumed to confer phenotypic resistance to the corresponding antibiotic. This approach is considered a rule-based WGS-AST method: the presence of antibiotic resistance determinant A in the genome confers resistance to antibiotic B. Therefore, rule-based methods rely on databases that contain information regarding the nucleotide or amino acid sequence of resistance determinants and which antibiotics they confer resistance to[240]. These methods are attractive as they offer the potential for results to be available in a matter of hours rather than days and to screen many more bacteria than culture-based methods. Moreover, rule-based approaches offer easily interpretable results and are currently the most popular methods for translation of WGS-AST into a clinical setting. However, the reliance on databases of known resistance determinants can introduce problems; as discussed in the previous chapter, these databases may introduce bias towards certain bacteria. This can lead to a "false negative" result: no genetic resistance determinant is present in an isolate, but the isolate is phenotypically resistant to a particular antibiotic. These rule-based methods therefore vary in their accuracy for antibiotic/taxon combinations[241]; to my knowledge, how accurate they are for gut bacteria has not yet been assessed.

In this chapter, I generate and study the *in vitro* phenotypes of a diverse set of human commensal gut bacteria in the HBC against nine clinically relevant antibiotics (all of which belong to classes on the WHO List of Essential Medicines). As previously discussed, phenotypic data can be used to determine breakpoints of antibiotic concentrations where isolates are considered susceptible or resistant. Since published breakpoints are limited for gut bacteria, I defined a system for determining whether the isolates I studied should be considered

susceptible or resistant. I then determined the enrichment of phenotypic resistance in isolates representing the four main gut phyla of bacteria: Actinobacteria, Bacteroidetes, Firmicutes and Proteobacteria[242]. I also combined this phenotypic data with the genotypic data from the previous chapter to determine whether a rule-based prediction method of antibiotic resistance inferred from the presence of known resistance determinants can be accurately applied to human commensal gut bacteria. I then applied a comparative genomics and phenotype approach to identify and study candidate novel antibiotic resistance genes.

4.2 Results

4.2.1 Phenotypic screening of antibiotic resistance in a subset of 73 HBC isolates

Having identified the presence of genetic antibiotic resistance determinants using CARD, I leveraged our ability to culture anaerobic gut bacteria[146,151] to assess the phenotypic response to antibiotics. 73 phylogenetically diverse isolates from the HBC (Fig. 4.1) were selected, representing all four phyla, approximately 10 % of the culture collection and 14 % of the number of different species (39/273) contained in the HBC: 16 isolates of Proteobacteria (seven species in two families); 21 isolates of Bacteroidetes (11 species in two families); 11 isolates of Actinobacteria (five species in two families); 25 isolates of Firmicutes (16 species in three families). These isolates have a total of 115 predicted resistance phenotypes against 16 of the 17 antibiotic resistance classes discussed in Chapter 4 (acridine dye was not included since it is not a clinically relevant antibiotic[243]). Each isolate was tested for *in vitro* sensitivity against nine antibiotics that are clinically relevant and on the WHO list of essential medicines[21]: amoxicillin and ceftriaxone (two different sub-types of beta-lactams), ciprofloxacin (a fluoroquinolone), erythromycin (a macrolide), gentamicin (an

aminoglycoside), metronidazole (a nitroimidazole), tetracycline, trimethoprim (a diaminopyrimidine), vancomycin (a glycopeptide).

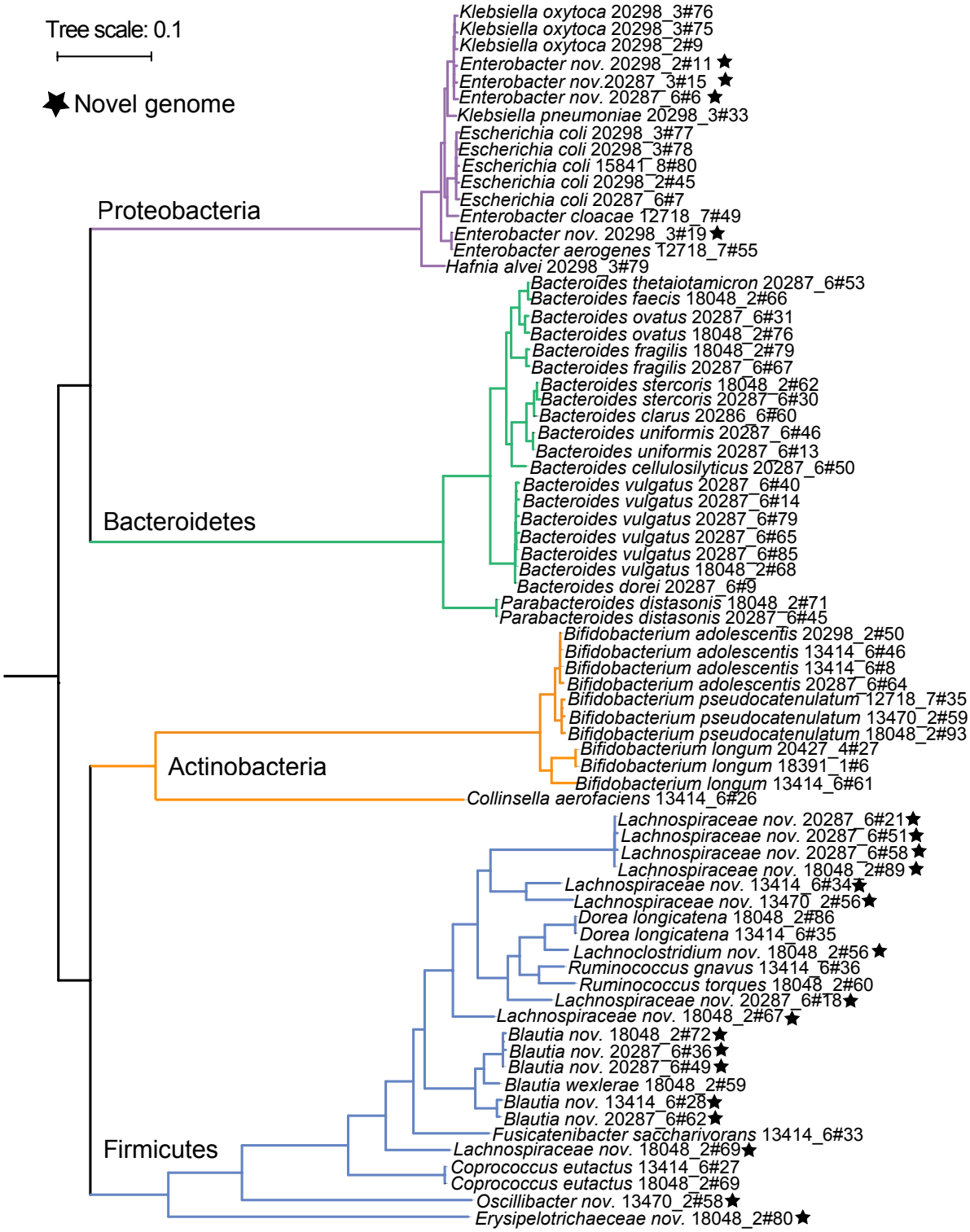

Figure 4.1. A phylogeny of 73 isolates from the HBC selected for selective phenotypic screening of antibiotic sensitivity. The phylogeny was inferred from 40 core genes of 73 whole genome sequences of gut microbiota isolated from healthy human faecal samples. These samples are a subset of the 737 isolates in the Human Gastrointestinal Bacterial Culture Collection (HBC) chosen to represent all four phyla.

To measure antibiotic sensitivity, I performed zone of inhibition analysis, due to the benefits described and because it is easy to perform in the confines of an anaerobic cabinet, where culturing anaerobic gut bacteria must take place. This method places single-concentration antibiotic disks (Table 4.1) on agar plates inoculated with the isolate of interest and measures the diameter of the zone of inhibition (the area surrounding the antibiotic disk where the isolate does not grow). An average zone of inhibition diameter was calculated from three biological replicates of each isolate/antibiotic combination (73 isolates x 9 antibiotics = 657 combinations). These phenotypic tests were set up by myself and Mr Mark Stares.

Table 4.1. Single-concentration antibiotic disks used for phenotypic sensitivity testing. Disks were manufactured by Oxoid and the concentration chosen was based on CLSI or EUCAST or BSAC guidelines as indicated in the table. Disk concentrations are usually selected based on guidance from EUCAST or CLSI breakpoint information; however, information for anaerobes is lacking. Therefore, concentrations were mainly selected based on advice for *Enterococcus*, as a Gram-positive, facultative anaerobic Firmicute. Though Bacteroides are Gram-negative, the same concentration disks were used for consistency and to allow comparisons.

Antibiotic	Antibiotic class	Disk Concentration (μg)	Based on guidelines
Amoxicillin	Beta-lactam: penicillin	10	*Enterococcus*; CLSI 2015
Ceftriaxone	Beta-lactam: cephalosporin	10	*Enterococcus*; EUCAST 2018
Ciprofloxacin	Fluoroquinolone	10	*Enterococcus*; CLSI 2015
Erythromycin	Macrolide	15	*Enterococcus*; CLSI 2015
Gentamicin	Aminoglycoside	30	*Enterococcus*; EUCAST 2018
Metronidazole	Nitroimidazole	5	Miscellaneous; BSAC 2015
Tetracycline	Tetracycline	30	*Enterococcus*; CLSI 2015
Trimethoprim	Diaminopyrimidine	5	*Enterococcus*; EUCAST 2018
Vancomycin	Glycopeptide	30	*Enterococcus*; CLSI 2015

Density curves were plotted for average zone of inhibition diameter for each antibiotic (Fig. 4.2), where each dot represents an average zone size for a particular isolate/antibiotic combination. These graphs therefore represent the proportions of isolates in each phylum with a certain zone size. Visualising the data in this way identifies trends in the range and frequency of zone of inhibition sizes for each antibiotic and each phylum. A larger zone of inhibition size (e.g. 60 mm) suggests that that isolate was very sensitive to that antibiotic; in contrast, a smaller zone of inhibition size (e.g. 5 mm) suggests that the isolate was less sensitive.

A very narrow range of zone of inhibition sizes occurs several times: Actinobacteria with amoxicillin, gentamicin, tetracycline and vancomycin; Bacteroidetes with gentamicin; Proteobacteria with ciprofloxacin, gentamicin, metronidazole, tetracycline and vancomycin (Fig. 4.2) These narrow distributions generally appear around the mid-range of average zone of inhibition sizes (20-40 mm), suggesting that all the isolates tested in those categories were sensitive (i.e., the antibiotic had an effect on these isolates). The exception is for Bacteroidetes and gentamicin, where the majority of isolates were not sensitive at all (zone of inhibition = 0 mm). Only two Bacteroidetes isolates were slightly sensitive with a zone of inhibition that was always smaller than 10 mm. This suggests that this phylum is particularly unaffected by gentamicin. Moreover, in some cases all isolates of a phylum were not sensitive to an antibiotic, such as Proteobacteria for metronidazole and vancomycin. Much bigger ranges in zone of inhibition size and thus sensitivity are also present: for example Actinobacteria with metronidazole; Bacteroidetes with amoxicillin, ceftriaxone, erythromycin and tetracycline; Firmicutes with ceftriaxone, erythromycin, tetracycline and trimethoprim. This indicates that some isolates in those phyla are more resistant to these antibiotics than other isolates. Since only eight bacterial families are represented by these 73 isolates, and the majority of isolates

in each phylum belong to a single family, the observed variation in antibiotic sensitivity also occurs within bacterial families as well as phyla.

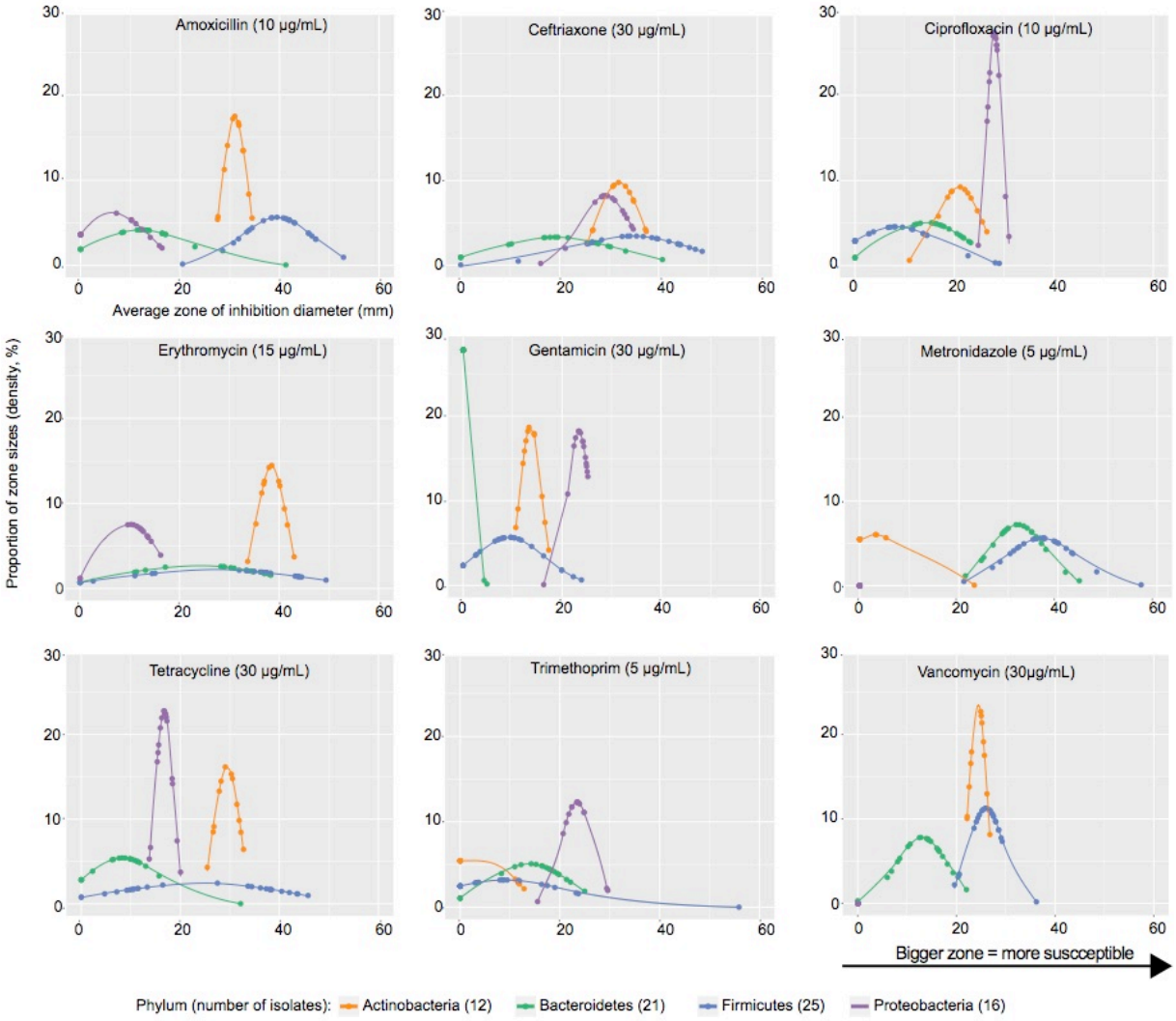

Figure 4.2. Density curves of zone of inhibition size among 73 isolates of the four main phyla of human gut bacteria. The isolates were screened for phenotypic antibiotic resistance against the following antibiotics: amoxicillin and ceftriaxone (beta-lactams), ciprofloxacin (a fluoroquinolone), erythromycin (a macrolide), gentamicin (an aminoglycoside), metronidazole (a nitroimidazole), tetracycline, trimethoprim (a diaminopyrimidine), and vancomycin (a glycopeptide) using disks containing a single concentration of an antibiotic. The screen was repeated in three biological replicates and zone of inhibition diameter averaged per isolate. Density curves for average zone of inhibition size was plotted for each phyla and antibiotic: each dot represents an average zone of inhibition size for a particular isolate/antibiotic combination. The x axis is the zone of inhibition size in millimetres: a larger average zone of inhibition means that isolate was more sensitive to an antibiotic than a smaller zone of inhibition. The y axis is the density or proportion of isolates with a particular zone size. Some phyla have a very narrow range of zone of inhibition sizes, e.g. Proteobacteria and ciprofloxacin or Actinobacteria and vancomycin. Other phyla have a much bigger range in zone of inhibition or susceptibility, e.g. Firmicutes and tetracycline. Mr Mark Stares helped generate zone of inhibition data.

4.2.2 Comparison of zone of inhibition sizes between isolates with and without genetic antibiotic resistance determinants

Combining the phenotypic data for these 73 isolates with the genomic data generated in the previous chapter it is possible to assess whether the antibiotic resistance genotype (or predicted phenotype) corresponds with the actual antibiotic resistance phenotype. Thus, the accuracy of the proposed rule-based method of predicting antibiotic resistance in gut bacteria is determined. To my knowledge, this is the first time this comparison of genotypic and phenotypic antibiotic resistance data has been performed for a diverse collection of gut bacteria isolated from healthy humans.

Initially, I used the Tukey method to study the range of zone sizes in isolates with and without predicted resistance and identify outliers with abnormally small zone diameters for each antibiotic (Fig. 4.3). This showed that there were isolates completely resistant (zone diameter = 0 mm) to an antibiotic despite not harbouring any resistance determinants from CARD to that drug. This occurs for all antibiotics and indicates that the CARD based predictions of genetic resistance determinants are not completely accurate for these gut bacteria. To investigate this further, the bacteria need to be categorised as resistant or susceptible to each antibiotic.

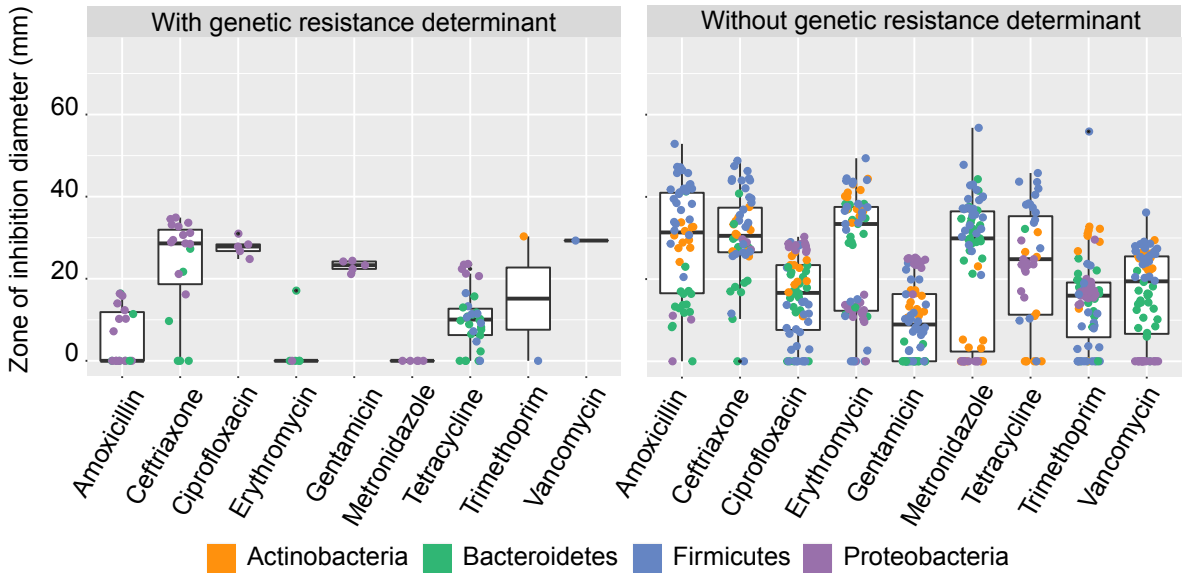

Figure 4.3. Range of average zone of inhibition between isolates with and without the presence of genetic resistance determinants. The average zone of inhibition diameters in isolates with predicted genetic resistance determinants against a certain antibiotic were compared against those in isolates with no predicted resistance. The boxplots show the interquartile range, median and limits at 1.5IQR above and below. Overall, there was a large range in zone of inhibition size in isolates without predicted genetic resistance determinants, with some isolates completely resistant to each antibiotic studied even in the absence of genomic resistance.

4.2.3 Defining a system for categorising gut bacteria as resistant or susceptible to antibiotics and considering the spectrum of antibiotics

Guidelines for defining resistance and susceptibility exist for a range of pathogens, though are less well-characterised in anaerobes and Gram-positive bacteria. Moreover, they are used for categorising closely related taxa, usually at the species level. In this thesis I wanted to compare at higher taxonomic levels, therefore I defined a scale using the phenotypic data I generated. Where very narrow ranges in sensitivity occur, cut-off points (breakpoints) in zone of inhibition size for categorising isolates as susceptible or resistant will be closer together (e.g. zone > 7.5 mm = susceptible but zone < 2.5 mm = resistant). Where broad ranges in sensitivity occur, breakpoints will be further apart (e.g. zone > 45 mm = susceptible and zone < 15 mm = resistant). This makes it clear that breakpoints for susceptibility/resistance must be defined for each antibiotic. Ideally, this would also be defined for individual taxa – in pathogens this is

done at the species level using thousands of clinical and sometimes environmental isolates. In this study, there are not enough isolates to be able to do that and a scale was defined using all 73 HBC isolates for each individual antibiotic. I determined the interquartile range of the zone of inhibition sizes and categorised isolates 'resistant' if the zone of inhibition size was in the lower quartile or smallest 25 % of all zone sizes for a particular antibiotic (Fig. 4.4, Table 4.2). Similarly, isolates were categorised 'susceptible' if the zone size was in the upper quartile or largest 25 % of all zone sizes for a particular antibiotic. Isolates with zone sizes in the middle 50 % require further testing to determine whether or not they should be considered susceptible or resistant and are referred to as intermediate sensitivity; these will be excluded from downstream analyses.

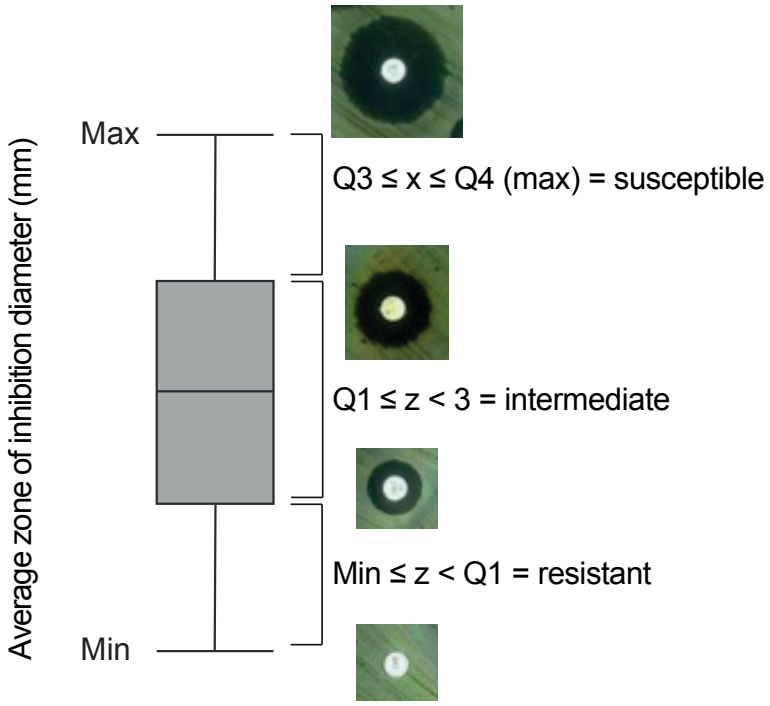

Figure 4.4. Proposed scale to define antibiotic resistance and susceptibility in human gut microbiota. I measured antibiotic susceptibility using single-concentration antibiotic disks and zone of inhibition size for nine antibiotics in 73 isolates (657 phenotypes in total). I ordered the zone of inhibition sizes for an individual antibiotic from largest to smallest and determined the interquartile range. I considered isolates 'resistant' if the zone size was in the lower quartile or smallest 25 % of all zone sizes for a particular antibiotic. Similarly, isolates were considered 'susceptible' if the zone size was in the upper quartile or largest 25 % of all zone sizes for a particular antibiotic. Isolates with zone sizes in the middle 50 % require further testing to determine whether or not they should be considered susceptible or resistant. Z = zone of inhibition diameter.

Table 4.2. Zone of inhibition limits used to categorise isolates of human gut bacteria as antibiotic-resistant or -susceptible. Zone of inhibition sizes for nine antibiotics in 73 isolates of human gut bacteria were determined. I ordered the zone of inhibition sizes for an individual antibiotic from largest to smallest and determined the interquartile range. I considered isolates 'resistant' if the zone size was in the lower quartile (min ≤ z < q1, where z = average of zone of inhibition in millimetres) or smallest 25 % of all zone sizes for a particular antibiotic. Similarly, isolates were considered 'susceptible' if the zone size was in the upper quartile (q3 ≤ z ≤ q4 (max)) or largest 25 % of all zone sizes for a particular antibiotic. Isolates with zone sizes in the middle 50 % (q1 ≤ z < q3) require further testing to determine whether or not they should be considered susceptible or resistant. For gentamicin, metronidazole and trimethoprim, all isolates with no zone of inhibition (0.00 mm) were considered resistant; any zone larger than 0.00 mm but smaller than the q3 value was considered intermediate.

Antibiotic	Resistant (min ≤ z < q1)	Intermediate (q1 ≤ z < q3)	Susceptible (q3 ≤ z ≤ q4 (max))
Amoxicillin	0.00 ≤ z < 11.09	11.09 ≤ z < 36.76	36.76≤ z ≤ 52.85
Ceftriaxone	0.00 ≤ z < 25.88	25.88 ≤ z < 34.90	34.90 ≤ z ≤ 48.85
Ciprofloxacin	0.00 ≤ z < 8.12	8.12 ≤ z < 25.61	25.61 ≤ z ≤ 30.98
Erythromycin	0.00 ≤ z < 11.35	11.35 ≤ z < 37.19	37.19 ≤ z ≤ 49.45
Gentamicin	0.00 ≤ z ≤ 0.00	0.00 < z < 19.92	19.92 ≤ z ≤ 25.21
Metronidazole	0.00 ≤ z ≤ 0.00	0.00 < z < 35.72	35.72 ≤ z ≤ 56.80
Tetracycline	0.00 ≤ z < 9.87	9.87 ≤ z < 30.31	30.31 ≤ z ≤ 45.80
Trimethoprim	0.00 ≤ z ≤ 0.00	0.00 < z < 21.32	21.32 ≤ z ≤ 55.91
Vancomycin	0.00 ≤ z < 6.85	6.85 ≤ z < 25.47	25.47 ≤ z ≤ 36.205

With this data it is possible to comment on the spectrum of antibiotic efficacy against diverse human commensal gut microbiota. For example, ceftriaxone appears a good example of a broad-spectrum antibiotic (i.e., an antibiotic that impacts several different bacterial taxa): all four phyla contain isolates that are very sensitive to this antibiotic (zone of inhibition > 25 mm, the midpoint in the range of zone of inhibitions measured). However, the proportions of isolates vary: for Proteobacteria, only one in 16 isolates (6.3 %) was susceptible to ceftriaxone, with one isolate (6.3 %) being resistant and the other 14 of intermediate sensitivity (87.5 %). In Bacteroidetes, one in 21 isolates (4.8 %) was susceptible to ceftriaxone, 12 isolates were resistant (57.1 %), and eight isolates were of intermediate sensitivity (38.1 %). In Actinobacteria, three in 12 isolates (25 %) were susceptible, one isolate was resistant (8.3 %) and eight isolates were of intermediate sensitivity (66.7 %). In Firmicutes, 14 in 25 isolates (56 %) were susceptible to ceftriaxone, four isolates were resistant (16 %) and seven were of intermediate sensitivity (28 %). On the other hand, gentamicin would be considered a more

narrow spectrum antibiotic (i.e., an antibiotic that impacts only specific taxa): no Bacteroidetes isolates were susceptible, but 19 were resistant (90.5 %; the other two isolates were of intermediate sensitivity (9.5 %)). In addition, no Actinobacteria were susceptible to gentamicin; all Actinobacteria isolates were of intermediate sensitivity. Four Firmicutes isolates were resistant to gentamicin (16 %) and another four were susceptible, with 17 isolates (68 %) of intermediate sensitivity. Overall, these observations suggest that sensitivity varies within bacterial families and phyla and that different antibiotics affect certain families and phyla to different extents.

4.2.4 Comparison of genomic predictions of antibiotic resistance with bacterial phenotypes and identification of unpredicted resistances

Four key genotype/phenotype combinations can be defined by comparing these two datsets (Table 4.3): Confirmed Resistance (genetic resistance and phenotypic resistance both observed); Confirmed Susceptibility (no genetic or phenotypic resistance); Unpredicted Susceptibility (genetic resistance predicted but phenotypically susceptible) and Unpredicted Resistance (no genetic resistance predicted but phenotypically resistant).

Table 4.3. Genotype/phenotype combinations of antibiotic sensitivity. Genotypes were determined by predicting the presence of antibiotic resistance genes and mutations described in CARD in the genomes of 73 isolates of human gut bacteria. Phenotypes were determined by zone of inhibition antibiotic susceptibility testing and categorized as resistant or susceptible. This produces four possible genotype/phenotype combinations.

		Presence of genetic resistance determinant (genotype)	
		Yes (Predicted Resistance)	No (Predicted Susceptibility)
Phenotype	Resistant	Confirmed Resistance	Unpredicted Resistance
	Susceptible	Unpredicted Susceptibility	Confirmed Susceptibility

The distribution of these genotype/phenotype combinations among the 73 isolates is visualized in Figure 4.5 against their core genome phylogeny. Each cell represents a genotype/phenotype combination for a particular isolate and antibiotic. Visualising the data in this way shows patterns that can be generalized to each phylum. For example, Confirmed Resistances appear most common in the Proteobacteria and Bacteroidetes, whereas Actinobacteria have the fewest Confirmed Resistances. In contrast, Firmicutes appear to have the most Confirmed Susceptible isolates. In addition, the Proteobacteria isolates have the most Unpredicted Susceptibility genotype/phenotype combinations. All phyla feature Unpredicted Resistances; in some cases, these occur in every isolate of a particular phylum. For example, Proteobacteria and vancomycin; Bacteroidetes and Gentamicin; Firmicutes and ciprofloxacin. These observations will now be explored in more detail.

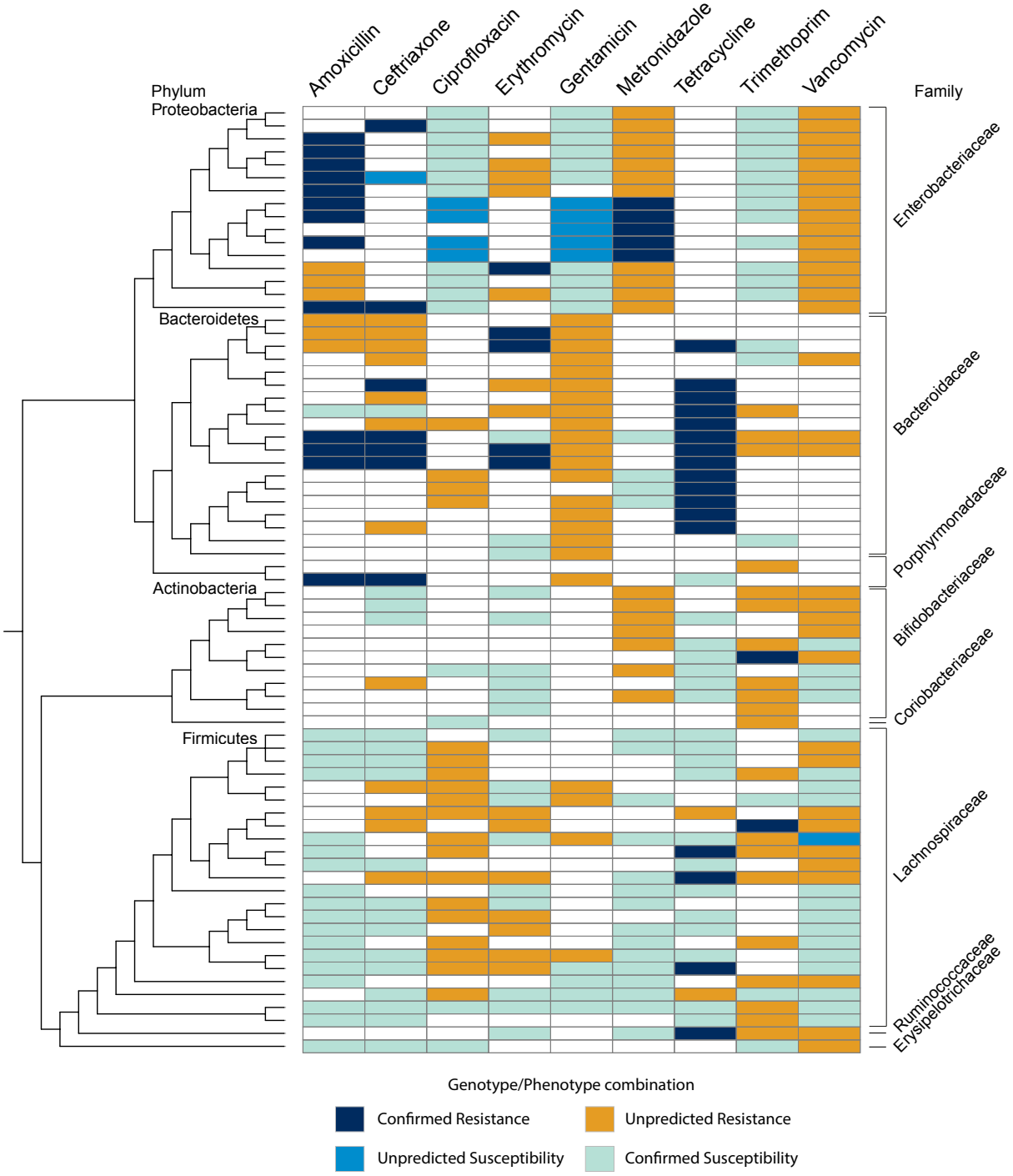

Figure 4.5. Distribution of antibiotic sensitivity genotype/phenotype combinations in 73 phylogenetically diverse isolates of human gut bacteria reveals many "unpredicted" resistances. The phylogeny was inferred from 40 core genes of 73 whole genome sequences of gut microbiota isolated from healthy human faecal samples. Each isolate was screened for antibiotic resistance to 9 antibiotics; in total 657 phenotypes were determined. Each cell in the figure represents an isolate's genotype/phenotype combination for a particular antibiotic. Dark blue = Confirmed Resistance (genetic resistance and phenotypic resistance both observed); Light blue = Confirmed Susceptibility (no genetic or phenotypic resistance); Mid-blue = Unpredicted Susceptibility (genetic resistance predicted, but phenotypically susceptible) and Unpredicted Resistance (no genetic resistance predicted, but phenotypically resistant). White cells represent combinations involving intermediate antibiotic sensitivity.

Examining the proportion of these genotype/phenotype combinations per phylum more closely (Fig. 4.6) showed that Confirmed Resistance is significantly higher in Bacteroidetes compared to the overall set of 73 isolates (q value < 0.0001; p-values determined by Fisher exact tests, adjusted using the Benjamini, Hochberg, and Yekutieli method for q-values, significant when q < 0.05) and Confirmed Susceptibility is significantly higher in Firmicutes (q value < 0.0001). Unpredicted Susceptibility mainly occurs in Proteobacteria, where it was significantly enriched (q value < 0.001), plus a very small amount in Firmicutes. All phyla demonstrate Unpredicted Resistance, but this occurs significantly more so than expected in Bacteroidetes (q value < 0.001). The proportion of Unpredicted Resistance genotype/phenotype combinations can be considered as the rate of False Negatives (i.e., the absence of CARD resistance determinants but phenotypic resistance indicates susceptibility was falsely predicted). In addition, the proportion of Unpredicted Susceptibility can be considered as the rate of False Positives (i.e., the presence of CARD resistance determinants but phenotypic susceptibility indicates resistance was falsely predicted). Thus, the overall False Negative rate for all genotype/phenotype combinations in all isolates was 38.6 %: the highest False Negative rate (Unpredicted Resistance, 51.9 %) occurs in Bacteroidetes, followed by 40.5 % in Actinobacteria, 39.6 % in Proteobacteria and 29.7 % in Firmicutes. The overall False Positive (Unpredicted Susceptibility) rate is 3.2 %: the highest False Positive rate occurs in Proteobacteria (11.0 %) and the lowest is 0.72 % in Firmicutes. There were no False Positives in Actinobacteria or Firmicutes. Therefore, antibiotic resistance seems to be more accurately predicted in some phyla of human gut microbiota than others.

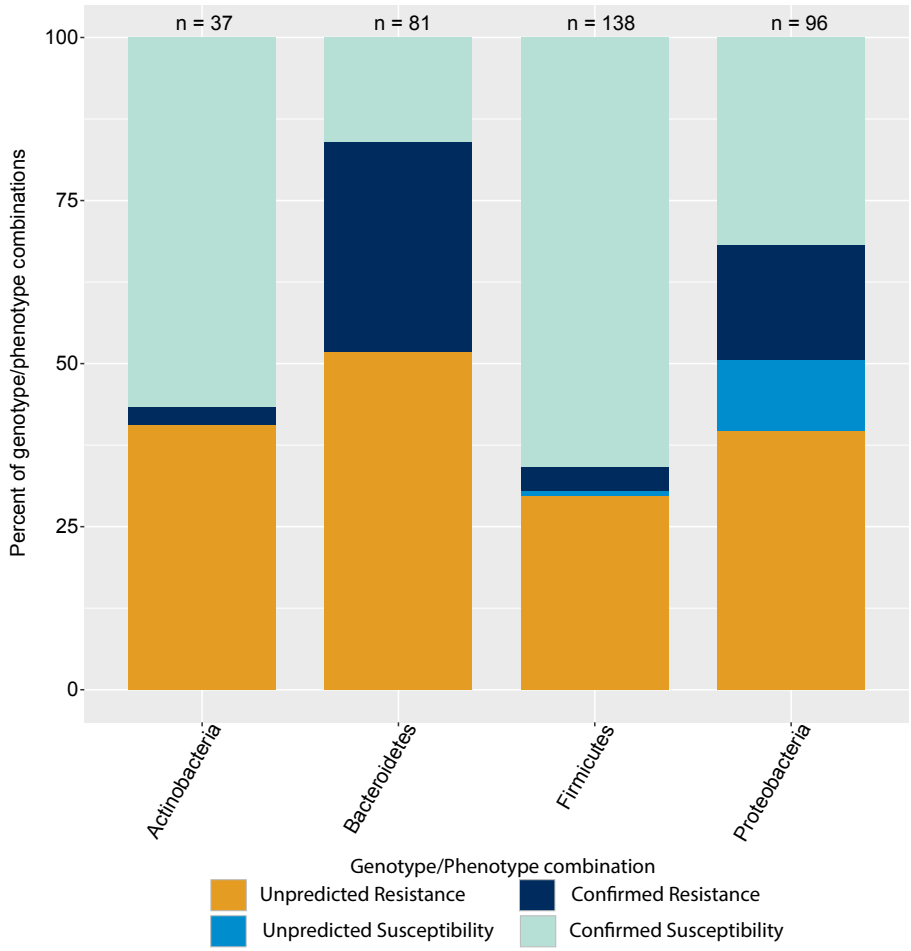

Figure 4.6. The proportion of genotype/phenotype combinations for each phylum. The number of genotype/phenotype combinations overall for all 73 isolates and each phylum was counted: All – 352; Actinobacteria – 37; Bacteroidetes – 81; Firmicutes – 138; Proteobacteria – 96. The proportion of specific combinations e.g. Confirmed Resistance, Confirmed Susceptibility, Unpredicted Resistance and Unpredicted Susceptibility was also determined across all 73 isolates tested and in each phylum. Confirmed Resistance was significantly enriched in Bacteroidetes (q value < 0.0001) compared to the overall Confirmed Resistance rate and was significantly lower in Firmicutes (q < 0.0001). Confirmed Susceptibility was enriched in Firmicutes (q < 0.0001) and occurred significantly less in Bacteroidetes than expected (q < 0.0001). Unpredicted Resistance was found in all phyla, but significantly more in Bacteroidetes (q = 0.0354). Unpredicted Susceptibility significantly occurs in Proteobacteria (q < 0.0001), with a very small amount in Firmicutes. P-values determined by Fisher exact tests, adjusted for multiple-testing using the Benjamini, Hochberg, and Yekutieli method; significant when q < 0.05. Unpredicted Resistance can also be considered a False Negative result and Unpredicted Susceptibility can be considered a False Positive result.

Assessing the proportion of genotype/phenotype for each antibiotic (Fig. 4.7) reveals that resistance is also more accurately predicted for some antibiotics than others. For example, both amoxicillin and tetracycline have significantly more Confirmed Resistances (34.2 % and 44.7 % respectively) than expected compared to the overall rate of Confirmed Resistances

(13.9 %) across all genotype/phenotype combinations (q values 0.035 and < 0.0001 respectively; p-values determined by Fisher exact tests, adjusted using the Benjamini, Hochberg, and Yekutieli method for q-values; significant when q < 0.05). Confirmed Susceptibility was not significantly different for any antibiotic compared to the overall rate (45.2 % of genotype/phenotype combinations across all antibiotics, ranging from 33.3 % to 50.0 % for individual antibiotics). Unpredicted Susceptibility is only observed for ceftriaxone, ciprofloxacin, gentamicin, and vancomycin, although was not significantly enriched in any of these antibiotics. The overall False Positive (Unpredicted Susceptibility) rate was 3.13 % of genotype/phenotype combinations, with the highest False Positive rate occurring for gentamicin (11.9 %), closely followed by ciprofloxacin (10.5 %), then dropping to 2.63 % for vancomycin; False Positives were not observed for amoxicillin erythromycin, metronidazole, tetracycline and trimethoprim. The overall False Negative (Unpredicted Resistance) rate was much higher, at 37.8 % of genotype/phenotype combinations. Whilst no antibiotic was enriched for more Unpredicted Resistances than expected, the False Negative rate was also highest for gentamicin (54.8 %), and similarly high for ciprofloxacin (50.0 %), vancomycin (50.0 %), and trimethoprim (48.8 %). The lowest False Negative rates occurred for amoxicillin (15.8 %) and tetracycline (5.26 %); indeed, False Negatives (Unpredicted Resistances) were observed significantly less often for tetracycline (q value < 0.0001) than expected, based on the overall rate. Overall, that all nine antibiotics have False Negative results (Unpredicted Resistances) further indicates that the isolates examined contain more antibiotic resistance than was predicted using the CARD database of known, clinically relevant antibiotic resistance determinants.

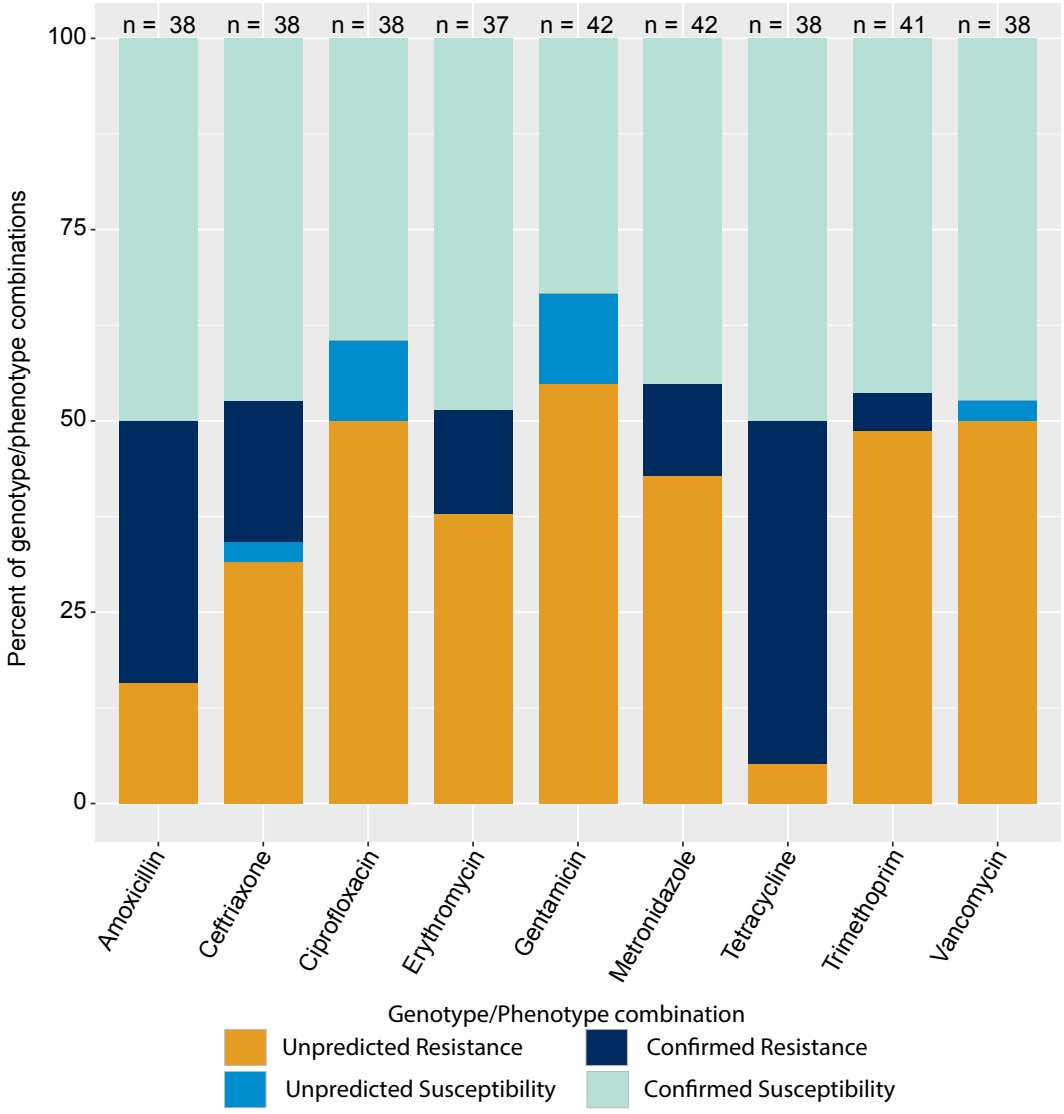

Figure 4.7. The proportion of genotype/phenotype combinations for each antibiotic. The number of genotype/phenotype combinations overall for all antibiotics ("All") and each antibiotic was counted: All – 352; Amoxicillin – 38; Ceftriaxone – 38; Ciprofloxacin– 38; Erythromycin – 37; Gentamicin – 42; Metronidazole – 42; Tetracycline – 38; Trimethoprim – 41; Vancomycin – 38. The proportion of specific combinations e.g. Confirmed Resistance, Confirmed Susceptibility, Unpredicted Resistance and Unpredicted Susceptibility was also determined across all and for each antibiotic. Confirmed Resistance is enriched in amoxicillin and tetracycline (q values = 0.035 and < 0.0001 respectively). Unpredicted resistance was observed significantly fewer times than expected for tetracycline compared to all antibiotics ($q < 0.0001$). P-values determined by Fisher exact tests, adjusted for multiple-testing using the Benjamini, Hochberg, and Yekutieli method; significant when $q < 0.05$. Unpredicted Resistance can also be considered a False Negative result and Unpredicted Susceptibility can be considered a False Positive result.

4.2.5 Comparison of antibiotic resistance databases and prediction methods

Having identified that the ARIBA with CARD rule-based method is not completely accurate, other databases and methods were applied to the 73 isolates to provide alternative predictions of resistance genotypes (Fig. 4.8): the CARD's own Resistance Gene Identifier tool (CARD-RGI) and ARIBA with the MegaRes, ResFinder, and SRST2-ARGANNOT databases were used. I then compared the newly generated resistance genotypes to the phenotypic data generated for the nine antibiotics to determine proportions of each genotype/phenotype combination (Confirmed Resistance, Confirmed Susceptibility, Unpredicted Resistance and Unpredicted Susceptibility) for each method. The proportion of each combination in the four new methods was compared to the proportion from the initial ARIBA with CARD analysis performed in this thesis. This shows that the ResFinder and SRST2-ARGANNOT databases had significantly higher proportions of Confirmed Susceptibility, but no significant differences in the other three genotype/phenotype combinations. All databases perform similarly in terms of the rate of False Negatives (Unpredicted Resistances), though it was highest when using ARIBA with the MegaRes database (45.5 %, compared to 45.1 % for ARIBA with CARD, 39.8 % for CARD-RGI, 39.2 % for ARIBA with ResFinder, and 38.9 % for ARIBA with SRST2-ARGANNOT). Despite a significant increase in Confirmed Resistance genotype/phenotype combinations with CARD-RGI, this tool also had a higher rate of False Positives (Unpredicted Susceptibility; 8.5 % compared to 3.1 % for ARIBA with CARD, 2.8 % for ARIBA with MegaRes, and 0.9 % for ARIBA with SRST2-ARGANNOT). Therefore, CARD-RGI in particular seems to overpredict resistance using genomic data (predicting resistance when the isolate is susceptible). For these reasons, the original ARIBA with CARD results are used for subsequent analyses.

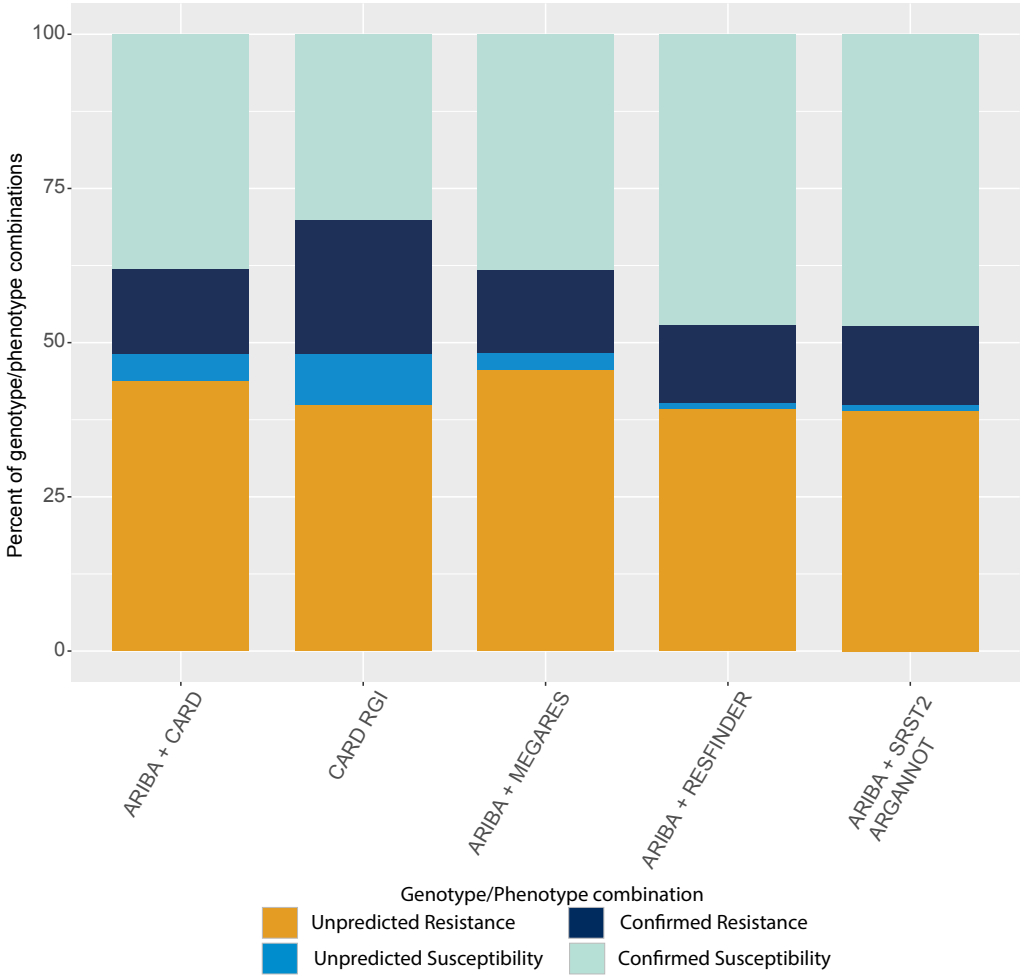

Figure 4.8. The proportion of genotype/phenotype combinations for each resistance database or method tested. The proportion of specific combinations e.g. Confirmed Resistance, Confirmed Susceptibility, Unpredicted Resistance and Unpredicted Susceptibility was determined for each database and method (total combinations = 352 for each method). These proportions were compared to the original ARIBA+CARD proportions using a two-proportion z test with Yates correction (q value significant when $q < 0.05$). The CARD Resistance Gene Identifier (CARD-RGI) method had significantly lower proportion of Confirmed Susceptibility ($q = 0.0385$) and significantly higher proportions of Confirmed Resistance and Unpredicted Susceptibility (q values 0.0103 and 0.0377 respectively). ARIBA with the RESFINDER and SRST2-ARGANNOT databases both had significantly higher proportions of Confirmed Susceptibility than ARIBA+CARD ($q = 0.0147$ and 0.0119 respectively). Unpredicted Resistance can also be considered a False Negative result and Unpredicted Susceptibility can be considered a False Positive result.

4.2.6 Identifying enrichment of unpredicted resistance to certain antibiotics in particular phyla

Determining which phyla have Unpredicted Resistances to certain antibiotics (Fig. 4.9) shows that Actinobacteria and Proteobacteria only have Unpredicted Resistance against three and four antibiotics respectively, whereas Firmicutes and Bacteroidetes demonstrate Unpredicted Resistance for six and eight antibiotics respectively. In particular, Actinobacteria were enriched in unpredicted metronidazole and trimethoprim resistance (q values 0.027 and 0.042 respectively; p-values determined by Fisher exact tests, adjusted using the Benjamini, Hochberg, and Yekutieli method for q-values, significant when $q < 0.05$). Bacteroidetes were especially enriched in unpredicted gentamicin resistance ($q < 0.0001$). Firmicutes were enriched in unpredicted ciprofloxacin resistance (q value 0.014), but fewer unpredicted metronidazole and vancomycin resistances were observed than expected (q values both 0.005). Proteobacteria were not enriched for any unpredicted resistances. Therefore, each phylum has different profiles of Unpredicted Resistances. In addition, this data supports the previous findings that Proteobacteria appears to have resistance more accurately predicted, and that unpredicted resistance overall was more common in non-Proteobacteria. However, as we previously saw that the Proteobacteria have large proportions of Unpredicted Susceptibility (Fig. 4.6), it appears that resistance may also be overpredicted in these isolates.

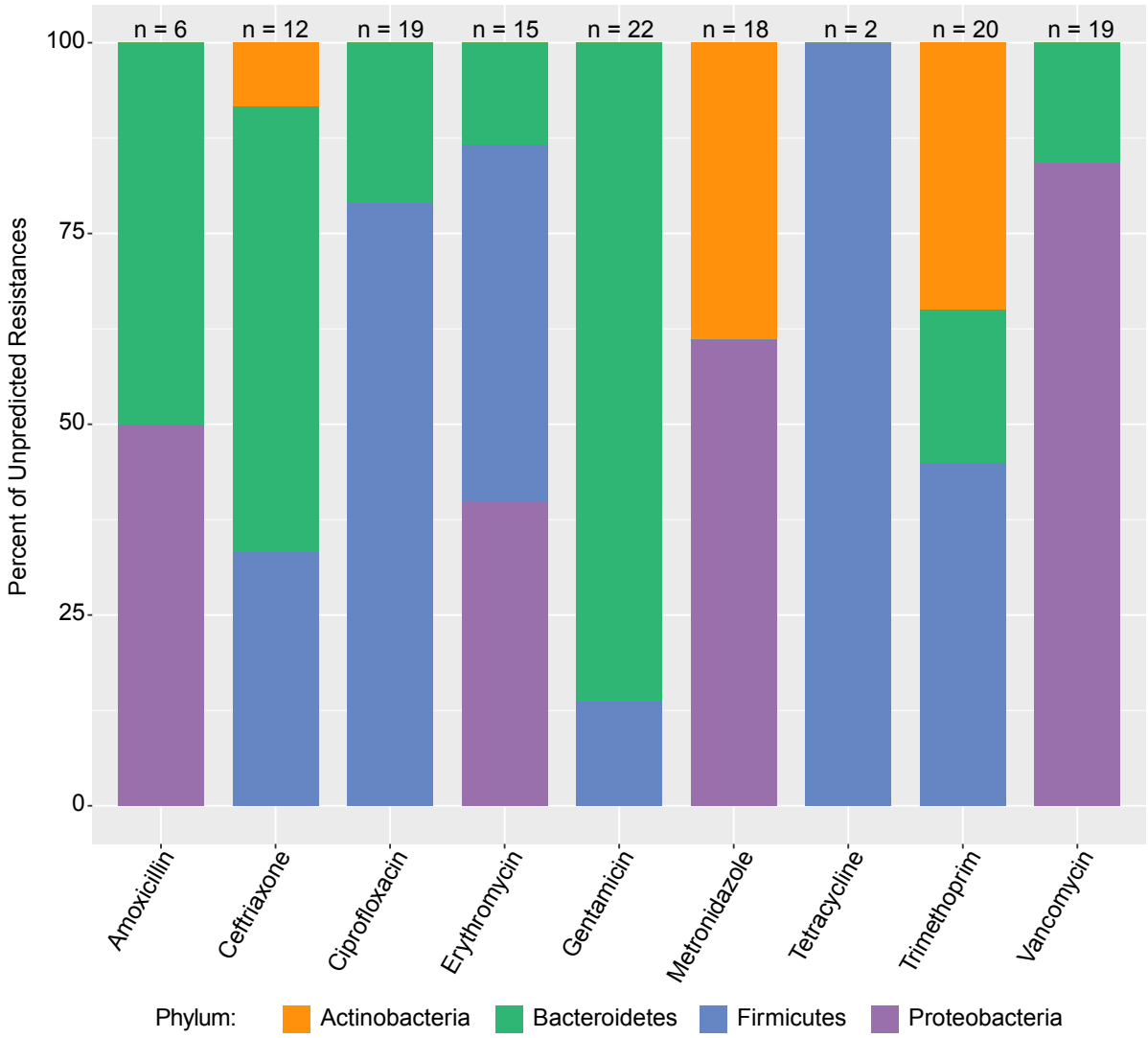

Figure 4.9. The percentage of Unpredicted Resistance antibiotic genotype/phenotype combinations by which phyla those combinations were observed in. Proteobacteria and Actinobacteria only demonstrate unpredicted resistance for three antibiotics, whereas Bacteroidetes and Firmicutes demonstrate unpredicted resistance for eight and six antibiotics respectively. Actinobacteria were enriched for Unpredicted Resistance to Metronidazole and Trimethoprim ($q = 0.027$ and 0.042 respectively). Bacteroidetes were enriched overall for Unpredicted Resistance ($q = 0.042$) but especially gentamicin Unpredicted Resistance ($q < 0.0001$). Firmicutes were enriched for ciprofloxacin Unpredicted Resistance ($q = 0.014$) but significantly fewer metronidazole and vancomycin Unpredicted Resistances were observed compared to the complete dataset (q values both 0.005). Proteobacteria were not enriched for any Unpredicted Resistances; significantly fewer Unpredicted Resistances were observed for ciprofloxacin ($q = 0.042$), gentamicin ($q = 0.027$), metronidazole ($q = 0.042$), trimethoprim ($q = 0.042$) and vancomycin ($q < 0.0001$). P-values determined by Fisher exact tests, adjusted for multiple-testing using the Benjamini, Hochberg, and Yekutieli method; significant when $q < 0.05$.

Considering the data so far, it is clear that the ARIBA + CARD method used here to initially predict antibiotic resistance in these human commensal gut bacteria is not accurate. In addition, the isolates studied contain more antibiotic resistance than is predicted using a rule-based method with the CARD database of known, clinically relevant antibiotic resistance determinants. In particular, the Bacteroidetes isolates were enriched for False Negatives (Unpredicted Resistance) and Proteobacteria isolates were enriched for False Positives (Unpredicted Susceptibility).

4.2.7 Further investigations of unpredicted resistance

Unpredicted Resistance genotype/phenotype combinations represent instances of a mismatch between two important methods for determining antibiotic sensitivity (culture-based- and WGS-AST). Some of these unpredicted resistance observations are likely explained by intrinsic resistance; for example, where they occur in all isolates of a particular phyla, including gentamicin in the Bacteroidetes isolates studied or vancomycin in the Proteobacteria (Fig. 4.5). However, where these unpredicted resistances occur in isolates that are closely related to isolates susceptible to that same antibiotic, these may be explained by genetic resistance determinants that are novel, or not described in CARD. This offers the opportunity to look for candidate novel antibiotic resistance genes or mutations in isolates with unpredicted resistance. In particular, there are instances of unpredicted ceftriaxone resistances observed in Bacteroidetes and Firmicutes where closely related isolates demonstrate Confirmed Susceptibility or Confirmed Resistance. As ceftriaxone resistance is often mediated by beta-lactamase enzymes these examples may indicate the presence of novel beta-lactamases in these human commensal gut microbiota.

Two particular isolates appeared good targets for novel beta-lactamases: *Bacteroides faecis* 18048_2#66 and *Lachnospiraceae nov.* 20287_6#18. These isolates both had ceftriaxone zone of inhibition sizes of 0 mm, indicating complete resistance to this beta-lactam antibiotic. They were also in the top five most ceftriaxone-resistant isolates (Fig. 4.10), but were the only two of those five without genetic determinants of beta-lactam resistance in their genomes.

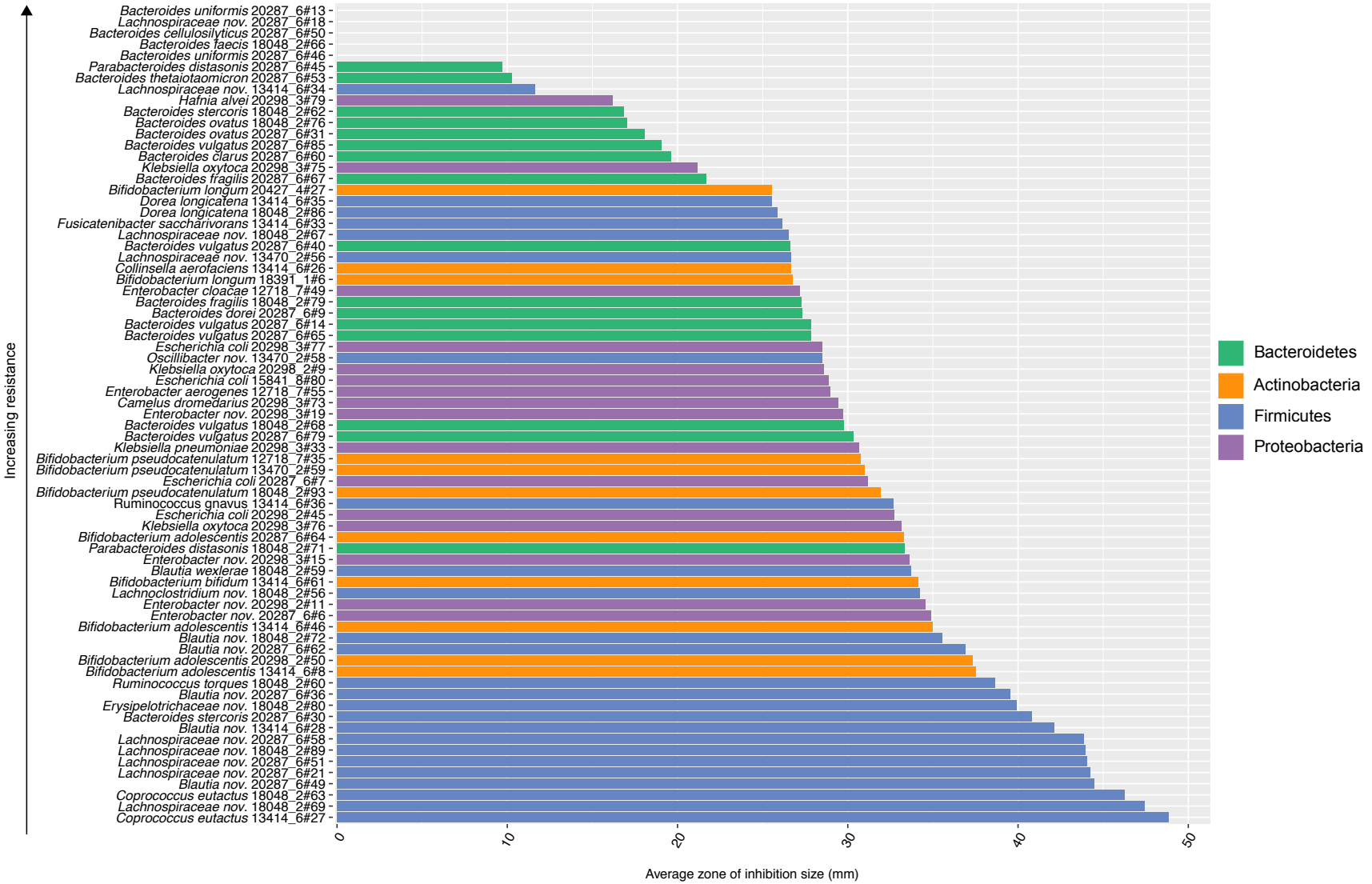

Figure 4.10. Ranking of isolates by ceftriaxone sensitivity. The average zone of inhibition sizes was used to rank the 73 HBC isolates by ceftriaxone sensitivity (bottom = most sensitive, largest zone of inhibition; top = least sensitive, smallest zone of inhibition). Five isolates were completely resistant to ceftriaxone (no zone at all).

4.2.8 Searching for novel antibiotic resistance determinants in human gut commensal microbiota

A comparative phenotyping and genomics approach was used to identify candidate novel resistance genes from *Bacteroides faecis* 18048_2#66 and *Lachnospiraceae nov.* 20287_6#18. I determined the average nucleotide identity (ANI), a measure of genomic similarity between the coding regions of two genomes, for each isolate under investigation and its closest two relatives from the HBC (Fig. 4.11). The two related *B. faecis* isolates, *B. faecis* 13470_2#65 and *B. faecis* 12718_7#26 were both more than 99 % similar by ANI to *B. faecis* 18048_2#66. *Lachnospiraceae nov.* 20287_6#18 had a very close relative of 99.24 % ANI, *Lachnospiraceae nov.* 8080_1#94, but the second next most closely related isolate from the HBC (*Coprococcus nov*. 20298_3#65) was only 81 % similar by ANI.

I determined the average Minimum Inhibitory Concentration (MIC) for ceftriaxone for each of the six isolates using Biomerieux Etests (antibiotic gradient strips) across three biological replicates (Figure 4.11). The average MIC for the two isolates with Unpredicted Resistance to ceftriaxone, *B. faecis* 18048_2#66 and *Lachnospiraceae nov.* 20287_6#18, was at least 256 µg/ml. The maximum concentration of ceftriaxone in the Etest strip was 256 µg/ml; these results mean that those two isolates were completely resistant to ceftriaxone at the maximum concentration tested and so the MIC is greater than or equal to 256 µg/ml. For *B. faecis* 18048_2#66, both related isolates *B. faecis* 13470_2#65 and *B. faecis* 12718_7#26 were more sensitive to ceftriaxone, with MICs under 60 µg/ml. The *Lachnospiraceae nov.* 8080_1#94 isolate shared the same phenotype as *Lachnospiraceae nov.* 20287_6#18 with an MIC of at least 256 µg/ml, and the *Coprococcus nov*. 20298_3#65 was much more sensitive with an MIC of just 0.5 µg/ml.

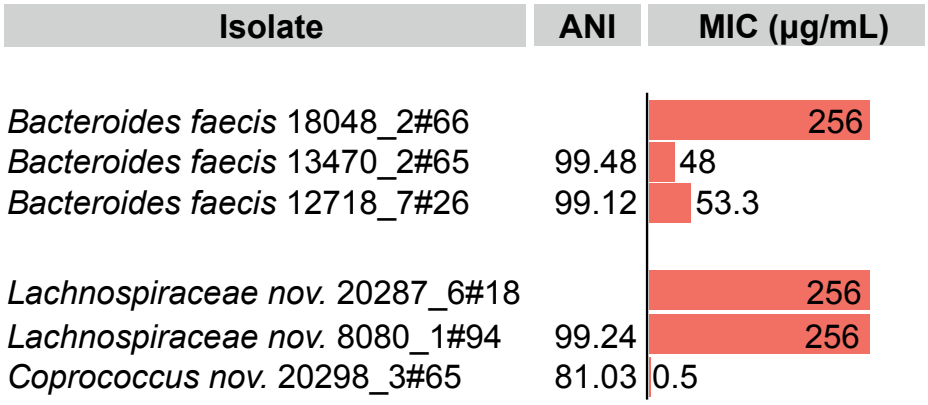

Figure 4.11. Comparison of ceftriaxone sensitivity in two sets of isolates from the HBC. *Bacteroides faecis* 18048_2#66 and *Lachnospiraceae nov.* 20287_6#18 were identified as having unpredicted ceftriaxone resistance. Their two closest relatives each were identified from the HBC using a 40 core gene phylogeny and Average Nucleotide Identity (ANI) was determined for the two closest relatives compared to the isolate in which unpredicted resistance was observed. The ceftriaxone Minimum Inhibitory Concentration (MIC, µg/ml) was determined for each isolate.

This data offers the opportunity to identify genomic differences between those isolates that may explain phenotypic differences. The *Lachnospiraceae nov.* 20287_6#18 isolate will be excluded since its most closely related isolate shared a ceftriaxone resistant phenotype and the ceftriaxone-susceptible *Coprococcus nov*. 20298_3#65 is a different species making genomic identification impractical. The *B. faecis* isolates, however, represent an ideal situation of very closely related isolates with differing phenotypic ceftriaxone sensitivity. In the rest of this section, I will investigate genomic differences between these *B. faecis* isolates to identify potential candidate beta-lactamase genes or mutations that may confer the unpredicted ceftriaxone resistance observed in *B. faecis* 18048_2#66.

Core genome analysis was performed using Roary[201] on the three *B. faecis* isolates and identified 3652 genes shared by all three isolates ("shared core genes") and 614 genes unique to the ceftriaxone resistant *B. faecis* 18048_2#66 ("resistant-unique genes", absent from the two more sensitive *B. faecis* isolates). ShortBRED[209] was used to reduce the amino acid sequences of 235,009 proteins containing the phrase "beta-lactamase" in their name from the

NCBI Protein database as of July 2018) into a database of reference amino acid markers. This includes beta-lactamase regulatory proteins so can possibly account for regulatory mutations too. These markers were used to search the translated amino acid sequences of the 614 resistant-unique genes and the 3652 shared core genes with 90 % identity. Other cut offs were tested (Table 4.4); however for subsequent analyses the 90 % cut off was used as a high level of similarity is typically required to infer functionality. Seven of the shared core genes were found to have amino acid sequences 90 % similar to reference amino acid markers from beta-lactamases (Table 4.5). A single gene labelled "Group 2384" was annotated as a candidate beta-lactamase.

Table 4.4. Numbers of resistant-unique and shared core genes with similarity to beta-lactamase markers in the human gut bacteria isolate *Bacteroides faecis* 18048_2#66. ShortBRED[194] was used to reduce the amino acid sequences of 235,009 proteins containing the phrase "beta-lactamase" in their name from NCBI Protein database (as of July 2018) into a database of reference amino acid markers. These markers were used to search the translated amino acid sequences of the 614 resistant-unique genes and the 3652 shared core genes with a variety of similarity cut offs.

Identity cut off (%)	No. of resistant-unique genes matched	No. of shared core genes matched
90	1	7
80	1	15
70	1	20
60	2	30
50	3	52
25	23	244

Table 4.5. A summary of candidate beta-lactamases that may explain an unpredicted ceftriaxone resistance phenotype observed in the human gut bacteria isolate *Bacteroides faecis* 18048_2#66. This isolate was phenotypically resistant to ceftriaxone in the absence of genetic determinants of beta-lactam resistance described in CARD. The closest two relatives from the HBC were identified and also phenotyped; they were both more sensitive to ceftriaxone. Roary core genome analysis was performed to identify genes unique to the resistant isolate ("resistant unique") and genes shared by all three isolates ("shared core"). These genes were searched for sequences with 90 % similarity to amino acid markers derived from 230,009 beta-lactamase related proteins in the NCBI Protein database. The table describes the genes that were found to contain markers of these proteins, what the genes were annotated as by Roary and the protein that the observed marker is derived from.

Category	Gene	Annotation	NCBI Beta-lactamase marker hit
Resistant unique	*Group 2384*	Beta-lactamase domain-containing protein	WP004329300 MULTISPECIES: MBL fold metallo-hydrolase [Bacteroidales]
Shared core	*ampG1*	Major Facilitator Superfamily	NP812531 AmpG protein, beta-lactamase induction signal transducer [*Bacteroides thetaiotaomicron* VPI-5482]
Shared core	*ampG2*	Signal transducer	NP809947 signal transducer [*Bacteroides thetaiotaomicron* VPI-5482]
Shared core	*blaR1*	Transcriptional regulator	WP010538315 MULTISPECIES: M56 family metallopeptidase [*Bacteroides*]
Shared core	*Group 106*	TonB	WP062695069 M56 family peptidase [*Bacteroides thetaiotaomicron*]
Shared core	*Group 3492*	Protein of unknown function (DUF2874)	WP062695288 hypothetical protein [*Bacteroides thetaiotaomicron*]; Putative beta-lactamase-inhibitor-like, PepSY-like; pfam11396
Shared core	*Group 4547*	Protein of unknown function (DUF2874)	WP008766859 hypothetical protein [*Bacteroides thetaiotaomicron*]; Putative beta-lactamase-inhibitor-like, PepSY-like; pfam11396
Shared core	*Group 6146*	Putative exported beta-lactamase protein	WP008769828 DUF302 domain-containing protein [*Bacteroides fragilis*]; Beta-lactamase; pfam00144

Subsequently, I investigated whether any of these candidate beta-lactamases might explain the unpredicted ceftriaxone resistance in *B. faecis* 18048_2#66. Firstly, I determined whether the candidate beta-lactamase gene Group 2384 unique to the resistant isolates corresponded with increased ceftriaxone MIC. To do this, I looked for the presence of Group 2384 in the complete set of HBC genomes with 100 % sequence length and nucleotide identity. This gene was identified in 16 other HBC isolates, all in the Bacteroidetes phylum. Five of these Bacteroidetes isolates were excluded from further analysis due to the presence of other genetic beta-lactam resistance determinants (identified in the analyses discussed in Chapter 3). For the 11 remaining Bacteroidetes isolates, I measured the ceftriaxone MICs for their closest relatives in the HBC (Fig. 4.12). In theory, if Group 2384 was responsible for the ceftriaxone-resistance phenotype in *B. faecis* 18048_2#66 and potentially other Bacteroidetes isolates, I would expect the presence of Group 2384 in an isolate's genome to correspond with a higher ceftriaxone MIC. This correlation was not observed, suggesting that this candidate beta-lactamase may not be responsible for this phenotype, or that it is not functional in the other eleven Bacteroidetes isolates.

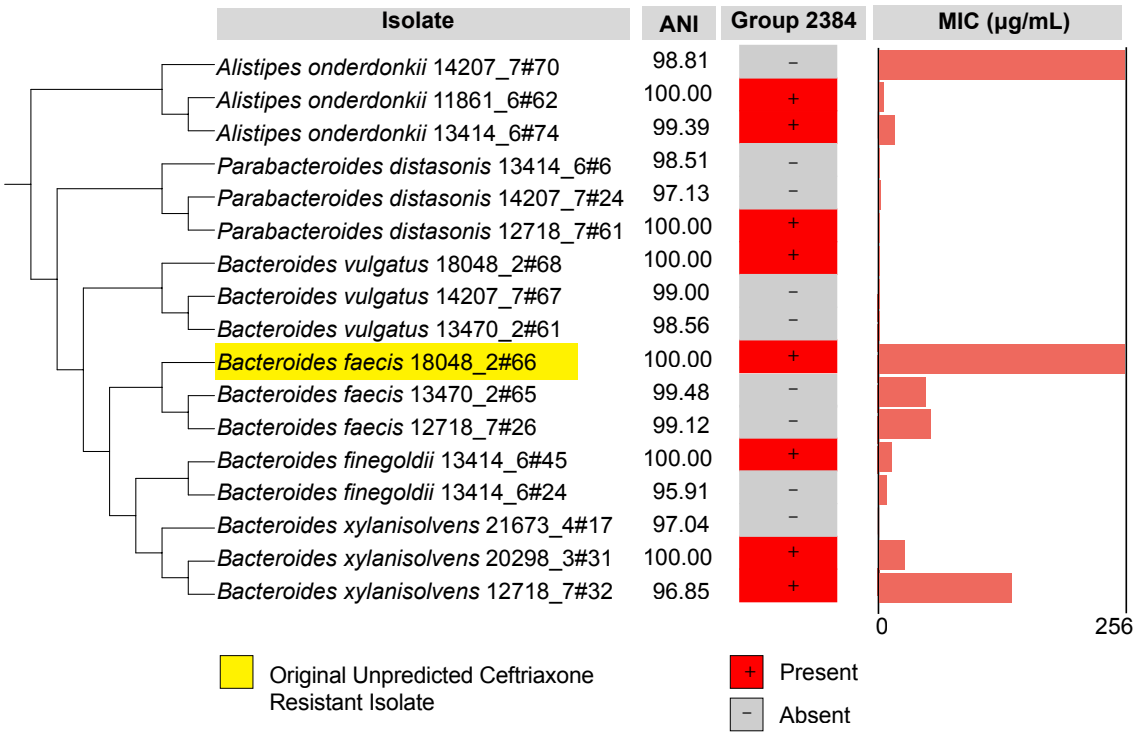

Figure 4.12. Ceftriaxone MIC in isolates with and without the Group 2384 candidate beta-lactamase gene. Group 2384 is a candidate beta-lactamase first identified in the Bacteroides faecis 18048_2#66, an isolate with unpredicted ceftriaxone resistance (highlighted in yellow). The presence of the Group 2384 gene was searched for in the HBC (100 % sequence length and ID) and was identified in 16 isolates. 11 Group 2384-positive isolates without any other predicted beta-lactam resistance (determined using CARD) plus each of their two closest relatives from the HBC were selected. ANI analysis was used to determine how similar the Group2384-negative isolates were to their closest Group 2384-positive relative. Ceftriaxone MICs were measured using Etests in three biological replicates of ach isolate. The presence of Group2384 was not correlated with a lower MIC for amoxicillin and ceftriaxone.

Although the presence of Group 2384 did not correlate with increased ceftriaxone MIC and thus resistance in other HBC isolates, this does not necessarily rule out its function as a ceftriaxone-resistance gene in *B. faecis* 18048_2#66. It is possible that it is not expressed in the other isolates or contains mutations that leave it non-functional. To rule out the latter hypothesis, I extracted and aligned the Group 2384 sequences from the twelve Group 2384-positive isolates I tested for phenotypic ceftriaxone sensitivity to infer a phylogenetic tree (Fig. 4.13). Whilst the Group 2384 genes were not identical, very few mutations were identified: *Bacteroides vulgatus* 18048 2#68 has base T at position 545 where the other eleven isolates have an A, seven of the Group 2384-positive isolates (including *B. faecis* 18048_2#66) have

base C at gene position 15, whereas the other five Group 2384-positive isolates have base T (Fig. 4.13). These mutations were all synonymous and did not alter the amino acid sequence. Therefore, it is unlikely that these mutations had any direct impact on the hypothetical function of Group 2384 as a ceftriaxone beta-lactamase, especially since six of the eleven Group 2384 sequences were identical to the one from *B. faecis* 18048_2#66.

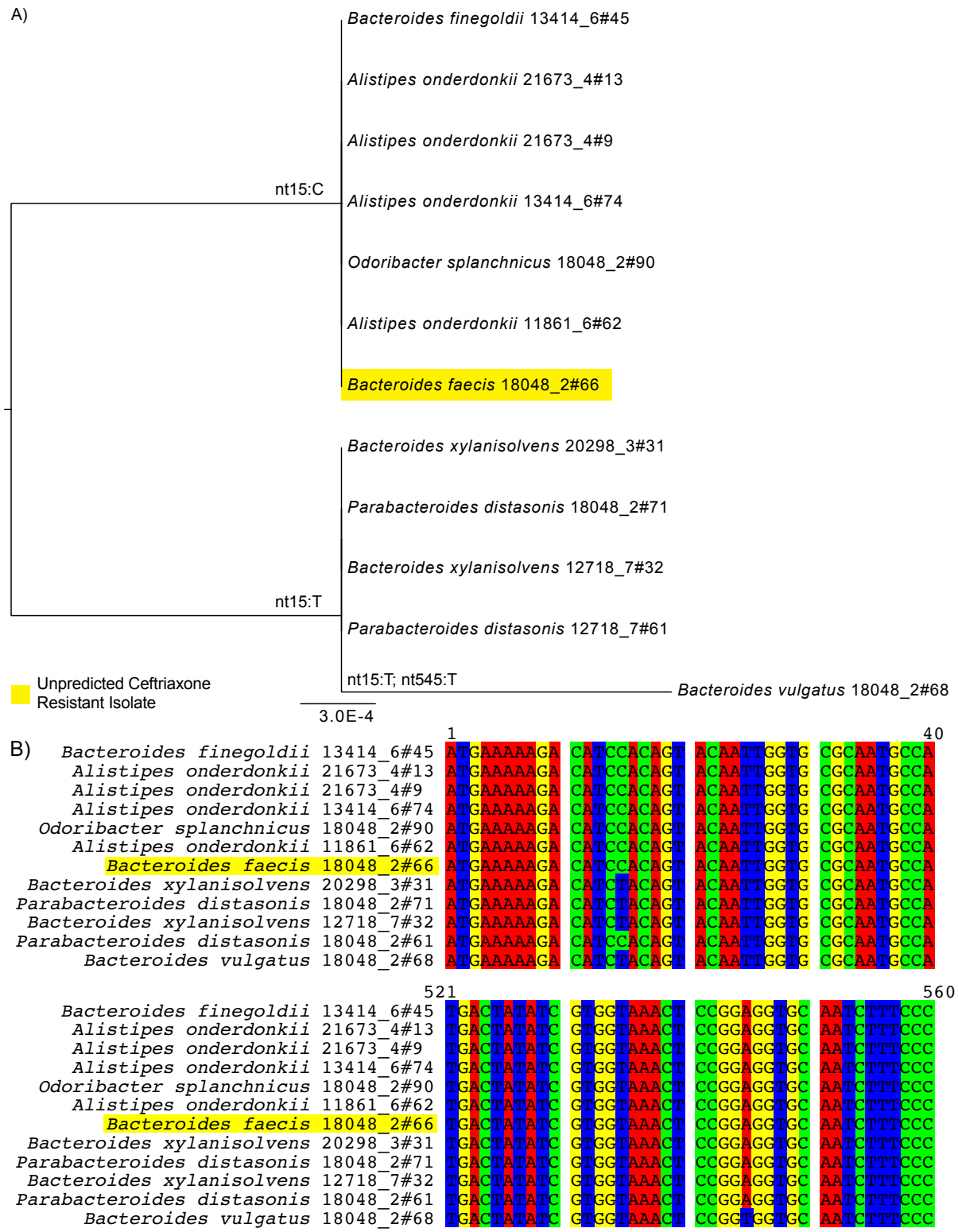

Figure 4.13. Group 2384 gene sequences from twelve HBC isolates. Group 2384 is a candidate beta-lactamase first identified in the *Bacteroides faecis* 18048_2#66, an isolate with unpredicted ceftriaxone resistance (highlighted in yellow). A: the nucleotide sequence for the Group 2384 gene was extracted from 11 other HBC isolates in which it was identified, aligned and used to infer a phylogeny and identify mutations (labelled at branch points). B: the alignment of nucleotides 1-40 and 521-560 are shown to illustrate the single nucleotide polymorphisms in the Group 2384 gene sequences.

To further investigate the potential function of Group 2384 as a ceftriaxone resistance-conferring beta-lactamase I performed a gain-of-function cloning experiment. A plasmid carrying a chloramphenicol resistance marker gene and the Group 2384 gene was designed with GeneArt (ThermoFisherScientific; Fig. 4.14). The Group 2384 gene was inserted within a tetracycline resistance gene, under the control of the tetracycline resistance gene promoter. The synthesised construct was transformed into electrocompetent *E. coli* cells via electroporation with the assistance of Mr Matthew Dorman. The transformed cells were grown on LB agar plates containing chloramphenicol (12.5 µg/mL) to check that the vector had been taken up by the *E. coli* cells. Transformed cells were also plated on LB agar containing chloramphenicol (12.5 µg/mL) and ceftriaxone at a concentration of 256 µg/mL (representing the observed *B. faecis* 18048_2#66 phenotype) and 4 µg/mL (slightly above the MIC of the untransformed *E. coli*). Whilst colonies were observed on the chloramphenicol control plates, no colonies were observed in the presence of either concentration of ceftriaxone. This suggests the Group 2384 gene may not infer ceftriaxone resistance and explain the unpredicted phenotype in *B. faecis* 18048_2#66.

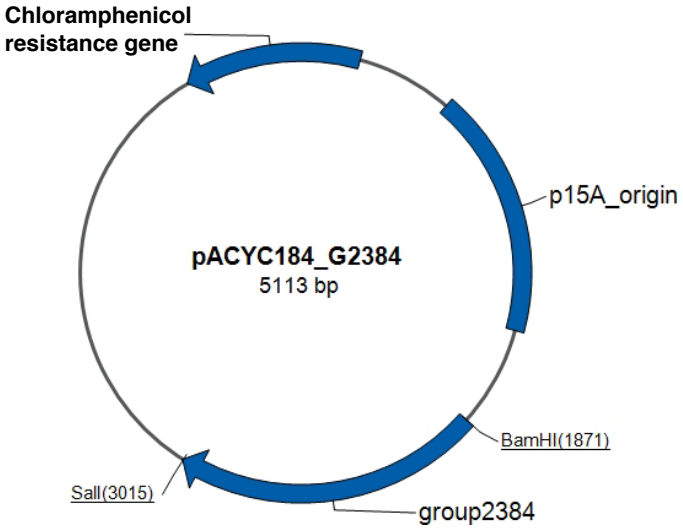

Figure 4.14. GeneArt construct containing Group 2384 candidate beta-lactamase gene. The plasmid pACYC184 was used as the cloning vector, designed and synthesised using GeneArt (ThermoFisherScientific) to contain the gene sequence of the Group 2384 candidate beta-lactamase identified from *Bacteroides faecis* 18048_2#66. The Group 2384 gene was inserted within a tetracycline resistance gene, under the control of the tetracycline resistance gene promoter. Mr Matthew Dorman assisted with the design of the plasmid construct.

It is possible that a mutation in a shared candidate beta-lactamase in the *B. faecis* isolates could cause the unpredicted ceftriaxone resistance phenotype. The gene sequence for each of the seven shared candidate beta-lactamases were extracted from the three *B. faecis* isolates and used to infer phylogenetic trees for each gene from the nucleotide sequences (Fig. 4.15). I looked for non-synonymous mutations that would alter the amino acid sequence of the protein product and could result in altered activity between the resistant *B. faecis* 18048_2#66 and more sensitive *B. faecis* 12718_7#26 and *B. faecis* 13470_2#65. Amino acid substitutions were identified in five of seven shared candidate beta-lactamases; in only one of these, "*ampG2*", was there a substitution unique to resistant *B. faecis* 18048_2#66. AmpG2 is a putative transporter related to AmpG, which possibly transports signal molecules into *P. aeruginosa* cells for the induction of the *ampC* beta-lactamase[244]. AmpG2 may function in a similar way; a non-synonymous mutation in this gene could potentially explain the phenotypic differences in these three isolates; however, it does not tell us exactly which beta-lactamase is degrading the ceftriaxone and amoxicillin antibiotics in *B. faecis* 18048_2#66. It could be controlling Group 2384, but further experiments would be required to investigate this.

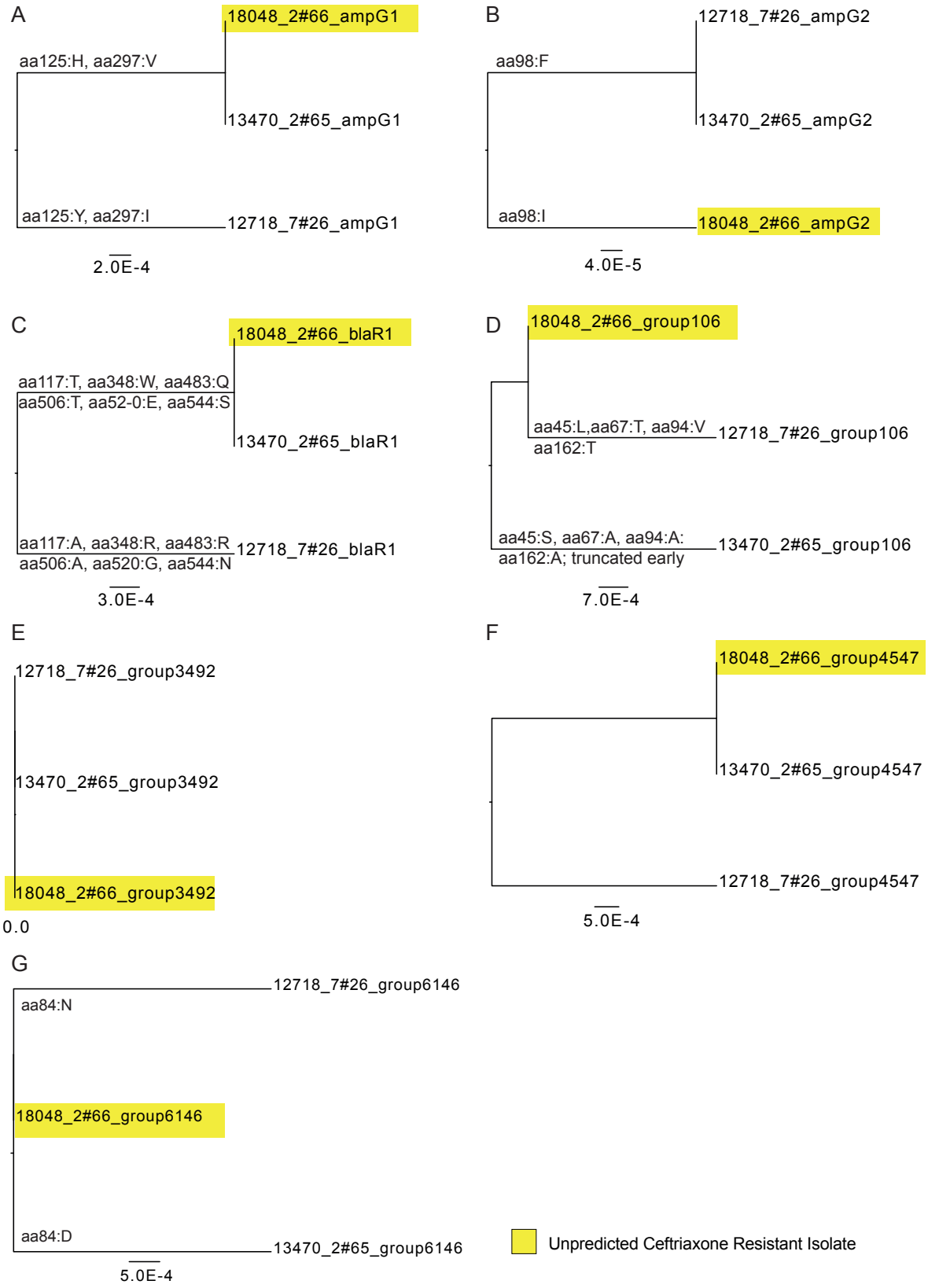

Figure 4.15. Gene phylogenies of candidate shared beta-lactamases in three HBC *Bacteroides faecis* isolates. Seven genes (A-G) shared between three isolates of *B. faecis* from the HBC were found to contain markers of beta-lactamase related proteins from the NCBI Protein database. One of the isolates was observed to be resistant to ceftriaxone in the absence of any known clinically relevant beta-lactam resistance determinant. Amino acid substitutions were observed in five of the seven genes; only in one gene was the mutation unique to the resistant isolate (B, ampG2).

Following these experiments, I applied a shotgun cloning approach to look for novel beta-lactam resistance genes from the whole genomes of *B. faecis* 18048_2#66 and *Lachnospiraceae nov.* 20287_6#18. Both isolates were regrown from HBC glycerol stocks kept at -80 °C, single colonies purified and grown in culture overnight. The full length 16S rRNA sequences were amplified from these cultures using 7f and 1510r PCR primers and sequenced by Sanger sequencing at Eurofins Scientific (Germany). The 16S rRNA sequences were then checked against the whole genome sequence data for these isolates to confirm their identity and check for contamination. Following this quality control, genomic DNA was extracted from the culture pellets of each isolate using phenol:chloroform extraction method by Mr Mark Stares. These genomic samples were used as starting material for the Copy Control Fosmid Cloning Kit (Lucigen) to clone 25-40 kb fragments into *E. coli* using fosmid vectors and Lambda phages. I determined the ceftriaxone MIC of the recipient *E. coli* strain using Etests as before.

The Copy Control Fosmid Cloning method was applied to each isolate individually and therefore represents shotgun cloning from purified isolates rather than mixed samples. This enables identification of the host of any novel antibiotic resistance genes discovered in these experiments. The kit includes control DNA of 40kb fragments, which was included as quality control alongside shotgun cloning of the two isolates with unpredicted ceftriaxone resistance. The fosmid vector carries a chloramphenicol resistance marker gene; the transformed cells were grown on LB agar plates containing chloramphenicol to check that the vector had been taken up by the *E. coli* cells. The transformed *E. coli* cells were also grown on LB agar plates containing ceftriaxone at a concentration of 256 µg/mL (representing the observed *B. faecis* 18048_2#66 phenotype) and 4 µg/mL (slightly above the MIC of the untransformed *E. coli*).

This procedure was repeated four times. Each time, hundreds of colonies were observed on the LB plates with chloramphenicol from the *E. coli* transformed with the control insert DNA.

However, only one or two colonies were observed from *E. coli* transformed with DNA from *Bacteroides faecis* 18048_2#66 or *Lachnospiraceae nov.* 20287_6#18 on the LB plates with chloramphenicol and no colonies on plates containing chloramphenicol and ceftriaxone.

4.3 Discussion

In this chapter I have determined the susceptibility or resistance of 73 phylogenetically diverse human commensal gut bacteria isolates against nine commonly used, clinically relevant antibiotics from the WHO list of essential medicines[21]. This phenotypic data helps to determine a comprehensive view of the impact of commonly used antibiotics across the diversity of gut microbiota. This offers insights into the spectrum of antibiotics which may be used to inform healthcare practices – such as which antibiotics to prescribe for infections caused by opportunistic pathogens from the gut. The main antibiotic from this study that would be useful to specifically target opportunistic anaerobic or gut bacteria is metronidazole, as most Bacteroidetes and Firmicutes tested were generally sensitive to this antibiotic, although this increases the likelihood of impacting more members of the gut microbiota.

In addition to knowing which antibiotics to use in the case of infections by opportunistic gut bacteria, this phenotypic data can advise on which antibiotics to avoid if trying to minimise the impact on commensal gut microbiota. For example, ceftriaxone is a broad-spectrum antibiotic[245] and therefore is useful for treating infections of unknown cause or with resistance to narrower spectrum antibiotics. Although typically administered via injection, not orally, ceftriaxone is known to have an impact on the commensal gut microbiota[246] and here I have showed that members of all four key gut microbiota could be affected. On the other hand, amoxicillin is also considered relatively broad spectrum, but here the only isolates that were

sensitive were Firmicutes, and one Bacteroidetes. Amoxicillin is one of the most commonly prescribed drugs in the world and typically administered orally for e.g. ear infections or throat infections. Amoxicillin also showed high levels of Confirmed Resistance; therefore, this antibiotic should also be used with caution as it may target Firmicutes, some of our most important gut microbiota. Moreover, there are already relatively high levels of amoxicillin resistance caused by known genetic determinants in commensal Proteobacteria that could potentially be transferred to pathogenic bacteria, especially under the selective pressure of amoxicillin therapy.

Another example of note is that of gentamicin: gentamicin is considered a broad-spectrum antibiotic, but one that does not work on anaerobes and streptococci. This is because gentamicin relies on oxygen-dependent transport into bacteria cells[24]. The results in Figure 4.5 demonstrate that this appears true for anaerobic Bacteroidetes, which were largely considered resistant to gentamicin. However, several Firmicute isolates were considered intermediate sensitivity or even susceptible, despite Lachnospiraceae (making up the majority of the Firmicutes isolates screened here) reported to be obligate anaerobes[247]. Therefore, gentamicin may have a more extensive impact on commensal gut microbiota than previously realized. Interestingly, there were no antibiotics that both Bacteroidetes and Firmicutes were both generally resistant to: all the antibiotics tested here have the potential to cause harm to common commensal gut microbiota.

The antibiotic resistance phenotypes were compared to the predicted resistance profiles based on the presence of genetic resistance determinants described in CARD, as determined in the previous chapter. The observations in this chapter follow a rule-based method of predicting antibiotic sensitivity phenotypes from antibiotic resistance genotypes. Rule-based methods have been found to be accurate for predicting antibiotic resistance in several species

of bacterial pathogens[240,242,248]. However, to my knowledge they have not been tested for accuracy in human gut commensal bacteria. I created a system to define the combined data as Confirmed Susceptibility, Confirmed Resistance, intermediate sensitivity (with or without genetic resistance determinants present), Unpredicted Susceptibility and Unpredicted Resistance. This revealed Unpredicted Susceptibility (False Positive results) to be most common in Proteobacteria and Unpredicted Resistance (False Negative results) most common in Bacteroidetes, but observed for all four phyla and all nine antibiotics.

By identifying unpredicted resistances, this suggests that the results in the previous chapter where antibiotic resistance determinants are enriched in Proteobacteria is not a true reflection of an enrichment of phenotypic antibiotic resistance. As in the previous chapter, it is important to note that database bias may explain this: a database designed from pathogenic bacteria (which are predominantly Proteobacteria, see Fig 1.3) may be more likely to identify similar antibiotic resistance genes or mutations in bacterial isolates more closely related to pathogens. It is likely that this partially explains the difference in observed enrichment of antibiotic resistance genotypes and antibiotic resistance phenotypes. Indeed, using alternative databases with generally similar False Positive and False Negative rates supports this. Moreover, the system I defined to classify isolates as phenotypically susceptible or resistant used all the isolates studied. In the future, the system should be redefined using isolates within more closely related taxa. Once studies approach the magnitude of those performed in pathogenic isolates (e.g. hundreds or thousands of isolates per species), the species level would be the most appropriate taxon to use. However, that resistance is better predicted to some antibiotics than others may also reflect the main mechanism of resistance for these enzymes: antibiotic resistance caused by the presence or absence of a particular gene (e.g. beta-lactamases or tetracycline resistance proteins are more likely to be called

accurately than a single nucleotide mutation). The largest proportion of the Unpredicted Resistances in Bacteroidetes were for gentamicin (46 %), which as discussed is thought not to work on anaerobic organisms such as Bacteroidetes. However, there are several other types of Unpredicted Resistance observed in the Bacteroidetes isolates, as well as relatively high levels of Confirmed Resistance. This indicates that Bacteroidetes make important contributions to the antibiotic resistance potential of the gut microbiota, which may have implications for the treatment of opportunistic pathogens caused by members of this phylum.

It was difficult to account for potential intrinsic resistances in this study to allow for the fact that current rules of antibiotic spectrum might not apply across the diversity of gut bacteria. I have already discussed one instance where Unpredicted Resistances might have been due to intrinsic resistance (gentamicin), but all the Proteobacteria isolates were also resistant to vancomycin as well. This antibiotic targets Gram-positive bacteria specifically, so this result is not surprising. In other cases, where there is variation on genotype/phenotype combinations between very closely related isolates, Unpredicted Resistances might indicate instances of novel antibiotic resistance genes or mutations.

In this study I identified two isolates with the highest possible ceftriaxone resistance measured in the absence of any genetic determinants of beta-lactam resistance. I investigated these isolates, *Bacteroides faecis* 18048_2#66 and *Lachnospiraceae nov.* 20287_6#18, for novel antibiotic resistance genes. This included detailed analysis and experiments regarding one particular candidate beta-lactamase (“Group 2384”), although I was not able to confirm its function. The Group 2384 gene was integrated into the pACYAC184 plasmid within a tetracycline resistance gene and thus was under the control of the tetracycline resistance gene promoter. Accordingly, the Group 2384 gene should have been expressed; however, the mRNA may not have been translated into the protein product with potential beta-lactamase

activity. It is also possible that the observed phenotypic variation in the three *B. faecis* isolates could be caused by differences in gene expression that cannot be detected with WGS alone. Q-PCR or RNA-sequencing of isolates growing in the presence of antibiotics, such as ceftriaxone, are two methods that may help determine if this is the case. In particular, qPCR experiments would help identify if the expression level of the candidate Group 2384 beta-lactamase is associated with the presence of mutations (as discussed towards the end of section 4.2.8) and/or differences in ceftriaxone sensitivity.

Furthermore, I was unable to identify any other candidate novel antibiotic resistance genes with the comparative genomics method. This could be due to the presence of novel beta-lactamases of less than 90 % similarity to beta-lactamases in the NCBI Protein database, though above 60 % similarity, no additional candidate beta-lactamases were identified. Below 50 % similarity, additional candidate beta-lactamase genes were identified and so these may be of interest; however, high sequence similarity is usually required to infer functional similarity[249]. This can be extended to other observations of Unpredicted Resistance; less stringent similarity cut offs when searching for the presence of antibiotic resistance determinants may produce more hits that could explain these observations, however, the Unpredicted Susceptibility rate is likely to rise in response.

The shotgun cloning method did not prove successful in identifying candidate novel antibiotic resistance genes during this study either. As hundreds of colonies were yielded from the control input DNA, this suggests the problem lies with the input DNA from *Bacteroides faecis* 18048_2#66 and *Lachnospiraceae nov.* 20287_6#18. The input DNA is therefore likely a highly critical factor for this protocol. In the future, this method will continue to be optimised within our laboratory as it would be a valuable tool to have available. Identifying genes conferring phenotypes such as antibiotic resistance and improve genome annotation in these organisms,

many of which are novel and/or uncharacterized, would be especially useful. In particular, any novel antibiotic resistance genes from gut bacteria should be curated into databases of known antibiotic resistance genes so that they can be included in antibiotic resistance surveillance programmes, such as the European Antimicrobial Resistance Genes Surveillance Network (EURGen-Net)[250].

Although susceptibility was generally better predicted than resistance, there were relatively high rates of Unpredicted Susceptibility in the Proteobacteria. This means that antibiotic resistance can be overpredicted. This also has relevance for healthcare: if a bacterium was predicted to be resistant to a particular antibiotic, a patient might be prescribed a different antibiotic that could pose more harm by being broader spectrum or more toxic than was required. Unpredicted susceptibility could be due to mis-calling a mutation as present when it is in fact absent, or by identifying the presence of housekeeping genes involved in the regulation of antibiotic resistance genes (such as the *vanR* regulator of vancomycin resistance[251]) . It could also be due to lack of expression or compensatory mutations to offset any fitness cost associated with the predicted antibiotic resistance mechanism. Alternatively, the observed genetic determinants of resistance could be taxon-specific. If a resistance determinant is observed in a taxon other than what it has been described in in the literature, something may be missing from the original host that is required for the determinant to be expressed or functional. It is again important to acknowledge the caveat of using CARD, whereby the majority of its antibiotic resistance genes and mutations have been described in just a few species of pathogenic bacteria. Therefore, in the context of commensal gut bacteria, they may not function as described in pathogenic isolates. It would be interesting to investigate these Unpredicted Susceptibilities in more depth. For example, the impact of individual genes on false predictions could be studied: if the presence of a gene or mutation

always results in Unpredicted Susceptibility (i.e., despite its presence the isolate is susceptible to that antibiotic), then determinants with this pattern could be filtered out as a poor predictor of antibiotic resistance. Taking this work forward must also involve looking for antibiotic resistance genes from other sources (e.g. curating those from functional metagenomic studies of bacterial communities) and determining specific relationships between the presence of antibiotic resistance determinants in a genome and antibiotic sensitivity. To really understand these relationships, more isolates of commensal gut bacteria should be studied (on a scale similar to that of pathogenic bacteria) and MICs should be determined for a more specific antibiotic sensitivity measurement.

This chapter highlights that currently a rule-based approach to estimating antibiotic sensitivity in human gut microbiota is not without flaws. Databases of antibiotic resistance determinants established through research on a relatively small number of pathogens, such as CARD, should therefore be used with caution when applied to more diverse, less well-characterised organisms – such as the human gut microbiota. As demonstrated, these methods can under-predict the antibiotic resistance of such isolates and show bias towards Proteobacteria. These databases can be useful and accurate for well-studied pathogenic bacteria[168,252], but if we are to accurately predict antibiotic resistance in the human gut microbiota and in metagenomic samples, more comprehensive databases of resistance genes are required. It may be necessary to have separate databases for common pathogens and for other bacteria, such as opportunistic pathogens, commensal gut microbiota, or other types of environmental bacteria. Since pathogens can acquire antibiotic resistance genes from environmental bacteria, these types of databases would be useful to help monitor the emergence of clinically relevant antibiotic resistance in clinical isolates of disease-causing bacteria. However, this will require additional similar studies comparing antibiotic resistance genotypes and phenotypes

but at a much larger scale, perhaps using high throughput alternative phenotyping methods such as plate based assays[253,254].

Moreover, functional metagenomics can be a very useful tool for identifying candidate novel antibiotic resistance genes[139,141,169,176,255,256], but requires optimization if used to study individual isolates. Putative ARGs are often annotated as such based on nucleotide or amino acid similarity to known antibiotic resistance genes/proteins and as such it is unknown whether they will confer phenotypic resistance. Determining the level of resistance that can be conferred and the distribution of novel resistance genes should become a routine part of these experiments. Since understanding antibiotic resistance genes among communities of bacteria and in individual, uncharacterized bacteria is difficult, perhaps prioritising bacteria/antibiotics of special interest – such as clinical or ecological relevance – is needed to focus the efforts of novel antibiotic resistance gene discovery.

Overall, this chapter shows that phenotypic antibiotic resistance in gut microbiota can vary between closely related isolates of commensal gut bacteria, much like in pathogenic bacterial species. Moreover, the Bacteroidetes and Firmicutes also demonstrate extensive phenotypic resistance, despite Proteobacteria appearing enriched for clinically relevant genetic determinants of antibiotic resistance in the previous chapter. These results further emphasise the role of the human gut microbiome as a reservoir for antibiotic resistance in terms of its occurrence and prevalence, but also that the extent of this is not yet fully known. In the next chapter, I will investigate the dynamics of antibiotic resistance in human gut microbiota and how commensal gut bacteria can evolve and spread antibiotic resistance.