

Chapter 7

Whole genome sequencing of ESBL *E. coli* carriage isolates

7.1 Chapter overview

This chapter describes the use of whole-genome sequencing (WGS) of ESBL producing *E. coli* to understand the drivers of gut mucosal ESBL-E carriage. I will begin with a description of the genomic landscape of the isolates from this study: starting with simple descriptions of *E. coli* phylogroup and multilocus sequence type (MLST) I will place the isolates from this study in the context of the *E. coli* population, followed by higher-resolution contextulaistaion using phylogenetics to place isolates from this study in the context of a global *E. coli* collection. I will describe the genetic basis of antimicrobial resistance in these isolates and explore the extent to which AMR genes tend to cluster together beyond what would be expected by chance. Finally, I will attempt to use the resolution offered by WGS to attempt to answer two specific questions: firstly, what is the mechanism of rapid increase in ESBL-E carriage prevalence following hospital admission and antimicrobial exposure we see in this study? Secondly, what is the likely unit of ESBL-E transmission in this study? Are bacteria, or mobile genetic elements (MGE) implicated? And if, MGE, which: plasmids, transposons, integrons - or a combination?

These questions, phrased in this way, seem difficult or impossible to answer given the available WGS data, but by slightly re-framing them they become tractable: first, what is the diversity of apparent hospital-acquired ESBL *E. coli* in comparison to apparent community-acquired isolates? Apparent hospital acquisitions could represent true acquisitions of, for example, a hospital-associated clone - but equally they could be an “unmasking” of minority variant *E. coli* in the microbiota, acquired in the community but not detected by culture because of

low abundance, until enriched for by antimicrobial exposure. If the diversity of apparently hospital acquired isolates is contained within the diversity of community isolates, this would lend support to this latter hypothesis. The second question - what is the unit of transmission in this system - can be re-framed by asking: what is the unit that is most conserved within patients, as compared to between patients? The questions then reduce to a dimensionality reduction problem: in order to address them both, it is necessary to classify either bacteria or MGE into mutually exclusive categories, in order to compare hospital to community isolates, and between-patient to within-patient. I describe the approach I have taken to this below.

7.2 Methods

7.2.1 Bioinformatic pipeline

The basic bioinformatic pipeline used is described in detail in Chapter 2, methods. Briefly, one *E. coli* colony from each patient sample was taken forward for DNA extraction and paired-end short-read whole genome sequencing using Illumina HiSeq X10 at the Wellcome Sanger Institute. Read quality control was undertaken with kraken v0.10.6 and braken v1.0 to assign reads to species[1] and WSI QC pipeline which maps a random 100 Mbases from each sample to a reference and calculates depth of coverage, number of heterogeneous SNPs, GC content and insert size. Samples that contained > 80% non *E. coli* reads were discarded and *de novo* assembly was undertaken with SPAdes v3.11.0[2]. Assembly statistics were calculated with QUAST v4.6.0[3] and completeness and contamination of the assemblies assessed by checkM v1.0.7[4]. Contaminated assemblies (with checkM-defined contamination of > 25%) or poor assemblies (with less than 1Mb assembled length) were discarded. Annotation was carried out with prokka v1.5[5] with a genus specific database from RefSeq and the roary v1.007 pan-genome pipeline[6] was used to identify a core genome. A core gene multiple sequence alignment was generated using maaft v7.205[7], SNP-sites identified using SNP-sites v2.4.1[8] and the resultant SNP alignment used to build a maximum likelihood phylogenetic tree using IQ-TREE v1.6.3[9], using ascertainment bias correction to correct for the fact that the input pseudosequence contained only variable sites, and using the ModelFinder module used to find the best fitting nucleotide substitution model. This calculates the likelihood of a number of different models and chooses the model with the lowest (best fitting) Bayesian Information Criterion, a statistic which penalises model parameters. Reliability of inferred branch partitions was assessed with 1000 bootstrap replicates. Trees were visualised in the ggtree v1.14.4 package[10] in R.

ARIBA v2.12.1[11] was used to identify AMR-associated genes using the SRST2 database[12],

to identify plasmid replicons using the PlasmidFinder database[13] and to perform *in silico* multi-locus sequence typing (MLST) using the database from <http://mlst.warwick.ac.uk/mlst/dbs/Ecoli> accessed via www.pubmlst.org. The β -lactamase genes *ampC1*, *ampC2* and *ampH* were excluded from the analysis of AMR determinants as they do not usually cause a resistant phenotype in *E. coli*. Because quinolone resistance often results from SNPs in the chromosome in the quinolone resistance determining regions (QRDRs) of the *gyrA*, *gyrB*, *parE* and *parC* genes - rather than acquisition of whole AMR-determining genes, as is the case with the other genes sought by ARIBA - these genes were downloaded from the comprehensive antimicrobial resistance database (CARD, <https://card.mcmaster.ca/>) and ARIBA used to call SNPs in them, with default settings. *E. coli* phylogrouping was performed with a quadruplex *in silico* PCR using the Clermont scheme[14] and *isPcr* v33x2 (<https://github.com/bowhan/kent/tree/master/src/isPcr>)

The rhierbaps v1.1.0 package in R[15] was used to cluster the core genome pseudosequence into sequence clusters (SCs). Two levels were used and these level 2 clusters used to test associations (see statistical analysis, below). To track putative mobile genetic elements ESBL-gene containing contigs were identified using BLASTn v2.7.0[16] of all contigs against the SRST2 database and then contigs containing any given ESBL gene were grouped by the ESBL gene they contained (for example, all *blaCTXM-15* gene-containing clusters were grouped together), and each group clustered using cd-hit v4.6[17] to produce mutually exclusive ESBL-gene-containing contig clusters for each identified ESBL gene. Henceforth, these clusters will be referred to as ESBL-clusters, for brevity. In order to attempt to determine the biological significance of the identified ESBL-clusters (i.e. what kind of MGE element they are likely to represent), basic statistics were plotted (number of samples contained within each cluster, length of longest contig in cluster in kbases, length distribution of all contigs is cluster relative to longest contig and distribution of sequence identity compared to the longest contig in the cluster). Presence of insertion sequences (i.e compound transposons), AMR determinants and plasmid replicons were identified by using BLAST with default settings of each ESBL-cluster representative sequence (as determined by cd-hit i.e one, the longest, for each ESBL-cluster) against the insertion sequence finder (ISfinder) database and the SRST2 database, filtering such that sequence identify was greater or equal to 95%, taking the top hit (as determined by bitscore) for any given location if there were two overlapping hits, and visualising the results in ggenes v0.3.2. To assess lineage association, the ESBL-clusters were mapped back to the core genome phylogeny.

7.2.2 Global *E. coli* collection

In order to place the isolates from this study in a global context, published *E. coli* assemblies were downloaded from the WSI servers. These included 149 ESBL-producing *E. coli* from a single centre study in Chachoengsao province, eastern Thailand[18]. In this study, human clinical isolates from standard care in Bhuddhasothorn hospital were selected on the basis of the ESBL phenotype, and environmental samples were collected as part of a cross sectional study and selectively cultured for ESBL-E in 2014-2015. I also downloaded assemblies of 362 enterotoxigenic *E. coli* (ETEC), selected for an ETEC genomic study from the Gothenburg University ETEC collection to represent a broad collection of ETEC isolated worldwide from 1980-2011[19]; 185 atypical enteropathogenic *E. coli* (aEPEC) sequenced for a study of aEPEC and selected from samples from the Global Enteric Multicentre Study (GEMS) in seven centres in Africa and Asia between 2007-2011[20]; and 94 *E. coli* from QECH in Blantyre, Malawi, a combination of invasive (bloodstream and CSF) and carriage isolates, selected for diversity in AMR phenotype from 1996-2014[21]. Details of the included samples are given in the appendix to this chapter.

Phylogroup and MLST were determined for these context genomes as described above. AMR genes were identified with Ariba and the SRST2 database, as above, and context genomes were classified as ESBL if they contained any Bush-Jacoby group 2be ESBL gene.

7.2.3 Statistical analysis

Ability of presence or absence of resistance determinants to predict phenotypic resistance as determined by antimicrobial sensitivity testing was expressed as sensitivity and specificity, with exact binomial confidence intervals. In order to explore clustering of AMR genes, the Jaccard index was calculated for a given AMR-gene pair using the jaccard v0.1.0 package in R. The Jaccard index, a measure of the similarity of two sets of data, is defined as *intersection over union*; in this context, for a given pair of AMR genes x and y , the Jaccard index $J(x, y)$ is the number of isolates that contain both gene x and y divided by the total number that contain either x or y . By definition it lies between 0 (x and y never co-occur) and 1 (x and y always co-occur). Co-occurrence matrices using the Jaccard index were plotted using the pheatmap v1.0.12 package in R. The statistical significance of co-occurrence of genes was assessed by generating 2x2 contingency tables for a given gene pair and p values generated using a Fisher's test with Bonferroni correction; a p value of less than 0.05 was considered statistically significant. Co-occurrence networks of genes occurring commonly together (defined as Jaccard index > 0.5) at a rate greater than expected by chance ($p < 0.05$ following Bonferroni correction) were plotted using igraph v1.2.2 and ggraph v1.0.2 in R.

To explore hospital or community associations of any given *E. coli* clade, the location of isolation was first plotted against the phylogenetic tree; location of isolation was classified as hospital, community, or recent hospital discharge (defined as a date of isolation within 2 weeks of hospital discharge). This latter category was used because it is possible that a patient could acquire an ESBL-E clone in hospital but only be sampled once leaving hospital; using only hospital isolated and community isolated categories could therefore introduce bias. Hospital or community association of each sequence cluster was assessed using a Fisher's exact test of proportion of hospital associated samples (defined as sum of hospital isolated and recent hospital discharge) for the given sequence cluster as compared to proportion of hospital associated samples in the remainder of the samples, with a Bonferroni correction for multiple comparisons. $p < 0.05$ was again considered statistically significant.

To compare within-patient to between-patient conservation of bacteria (as represented by core genome alignment and sequence cluster) and ESBL-containing MGE (as represented by the ESBL-clusters) several approaches were taken. Firstly, I assessed whether either sequence cluster or ESBL-cluster were conserved within an individual at all. I hypothesised that any within-patient correlation is likely to be a function of time: samples closer together in time may be more likely to be similar. To assess if this was the case for bacteria, pairwise core genome pseudosequence SNP distance was calculated using snp-dists v0.4 (<https://github.com/tseemann/snp-dists>) for all samples and plotted against the time difference (in days) between samples, within and between patients, and with a smoothed curve fitted using a general additive model with cubic splines. Because of significant overplotting, this was also plotted as a 2D density plot. Based on these plots, the within and between patient SNP distances were compared in two post-hoc defined groups binned by time distance between the samples (50 days or less vs. more than 50 days), and distributions compared with Kruskal-Wallace tests.

I then compared the within patient temporal clustering of ESBL-clusters and sequence clusters, by estimating the proportion of within-patient samples that contain the same ESBL-cluster or sequence cluster, as a function of time; essentially a temporal auto-correlation function. To estimate this, I considered pairwise comparison of all within-patient samples. For any given time between samples, t I defined a window of $+/-5$ days and estimated the probabilities as the number of all within-patient sample pairs in the window $[t - 5, t + 5]$ that contained the same sequence cluster or ESBL-cluster divided by the total number of all within-patient sample pairs within that time window. Exact binomial confidence intervals for these proportions were generated and probabilities plotted as a function of time. In order to estimate the probability of two samples containing the same sequence cluster or contig-cluster purely by chance, 1000 sample pairs were randomly drawn from all samples with replacement and the proportion of these samples that contained the same sequence cluster or ESBL-cluster calculated.

Finally, to inform the question as to what the likely unit of transmission in this system is, I assessed what was most conserved within patients, in pairwise sample comparison: bacteria (as represented by core gene sequence cluster), ESBL-containing MGE (as represented by ESBL-cluster), or both. Simple proportions in all-against-all pairwise comparison - stratified by whether between-patient or within-patient - were calculated: the proportion of samples that contain the same core gene sequence cluster only, the proportion of samples contain the same ESBL-cluster only, and the proportion that contain both sequence cluster and ESBL-cluster. Proportions were compared between within and between-patient strata in these three groups using Fisher's exact test, with $p < 0.05$ considered statistically significant.

7.3 Results

7.3.1 Samples and quality control

In total, 519 *E. coli* underwent DNA extraction and were shipped from Malawi to WSI; these represented all sequential isolates at the time of final DNA extraction, which occurred in two batches in February 2018 and October 2018. Kracken/Bracken read assignment of these samples is shown in Figure 7.1. The majority of samples have $> 90\%$ of reads assigned to *E. coli*; a minority have $< 90\%$ of reads assigned to *E. coli* but a very closely related species such as *Shigella*, and as such are likely to be pure *E. coli* culture with read misclassification. However, 12 samples have $> 80\%$ reads assigned to a non- *E. coli* species such as *Klebsiella pneumoniae*. These samples were assumed to represent upstream species misidentification or, perhaps more likely, selection of the wrong sample from the freezer archive for culture and DNA extraction, given that for any sample ID there are often several bacterial species identified and cryopreserved. These samples were excluded from further analysis.

Of the remaining 507 samples, there were a median (IQR) of 2339594 (2112842.5-2533930.5) reads, with a median (IQR) depth of coverage (obtained by mapping a random 100Mbases to a reference *E. coli* genome, Escherichia coli strain K-12 substrain MG1655, NCBI reference NC_000913.3) of 58 (51-66). One sample had an order of magnitude lower number of reads (291556) with depth of coverage 0; this was assumed to represent sequencing failure and it was excluded from further analysis.

The output from quast and checkM are shown in Figure 7.2, where N50 (the minimum contig length upon which at least half assembled bases are contained) is plotted as a function of total assembled length. The expected *E. coli* genome length is around 4.6Mb and most samples cluster close to this at a total assembled length of $\sim 5\text{Mb}$. However it is clear that some assemblies have failed, with low N50 and low assembled length. It is also apparent that some

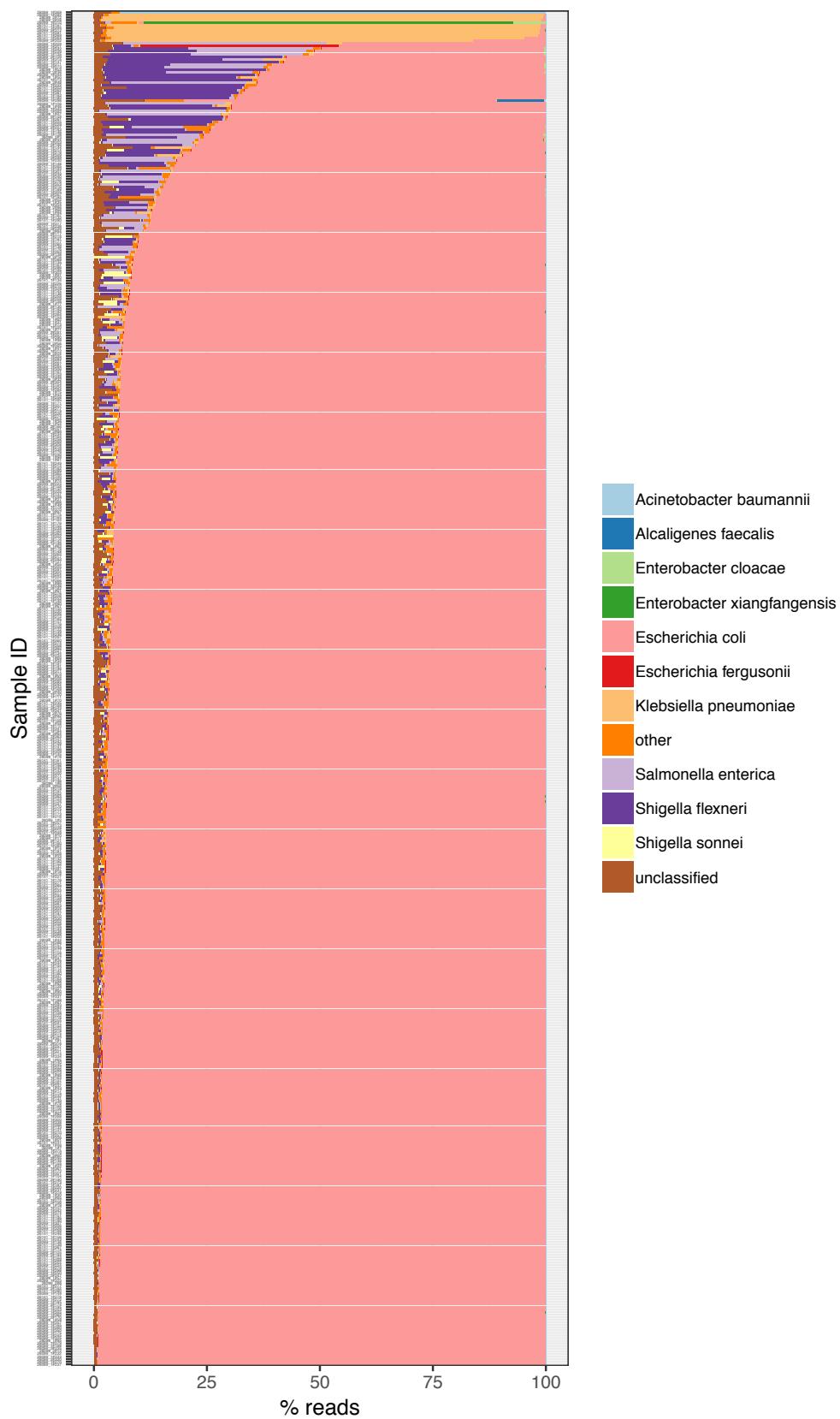


Figure 7.1: Species read assignment of all samples

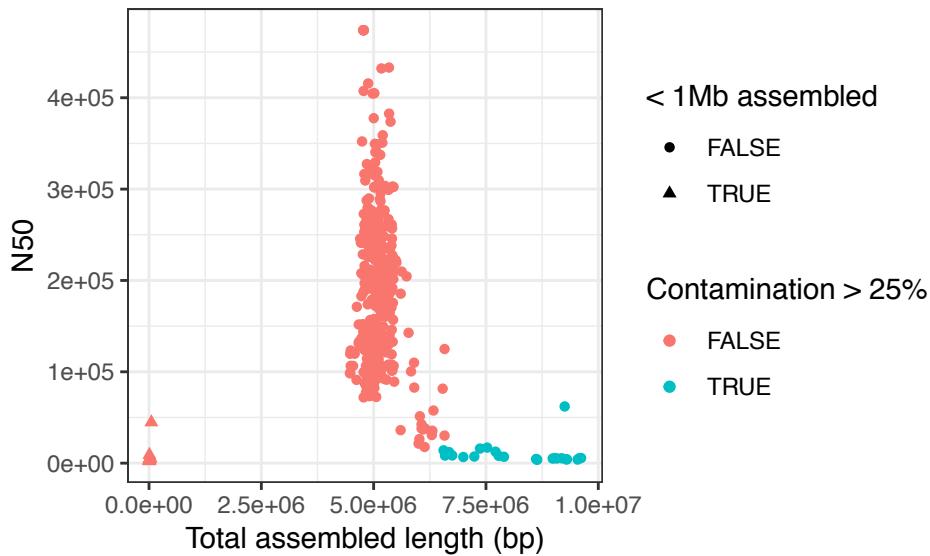


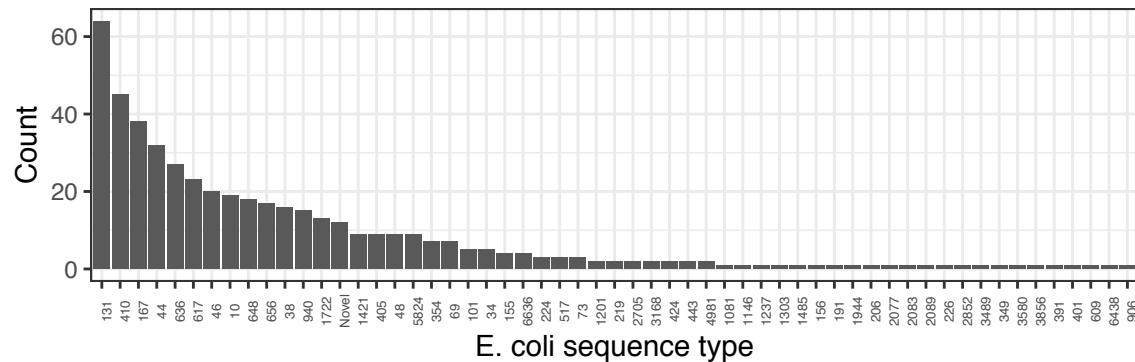
Figure 7.2: N50 as a function of total assembled length. Failed assemblies with less than 1Mb assembled shown as triangles. Contaminated assemblies with checkM-defined contamination above 25% shown in blue.

samples seem to be contaminated, as indicated by low N50 and much longer than expected total assembled length. Defining assembly failure as < 1Mb assembled length (triangles in the plot, n = 9) and contamination as checkM-defined contamination of > 25% (blue points in the plot, n = 24) and excluding both groups results in 33 further samples being excluded from further analysis.

In total, therefore, 46/519 (9%) of samples which were submitted for sequencing were excluded from downstream analysis. The remaining 473 samples represent 69% (474/686) of the cultured *E. coli* in this study, and were recovered from 230 participants. 354 are from patients with sepsis, 86 are from hospitalised inpatients and 33 are from community members, with a median of 2 (range 1-5) samples per participant. N50, total assembled length and number of assembled contigs are shown in the appendix to this chapter.

7.3.2 Phylogroup, MLST and core genome phylogeny of study isolates

The commonest *E. coli* phylogroup was phylogroup A: 204/473 (43%) samples belonged to phylogroup A, followed by phylogroup B2 (96/473 [20%]), F (53/473 [11%]), B1 (43/473 [9%]) and C (43/473 [9%]) and D (26/473 [5%]). Two samples were Clade I or II (so called cryptic clades) and 6/473 (1%) were unknown phylogroup using the Clermont PCR scheme. In the MLST analysis, 56 recognised sequence types (STs) were identified, and 12 samples were novel STs; however over half (249/473 [53%]) of samples were represented by the top seven most

Figure 7.3: *E. coli* multilocus sequence type distribution

frequent STs (Figure 7.3). ST131 was the most commonly isolated sequence type (64/473 [14%] of isolates) followed by ST410 (45/473 [10%] of isolates) and ST167 (38/473 [8%] of isolates).

The Roary pan-genome pipeline identified a core genome in the study isolates of 2966 genes, with a pan-genome of 26840 genes. The resultant core gene pseudosequence of length 1388742 bases contained 99693 variable sites, which were used to infer the maximum likelihood phylogenetic tree. The IQTREE ModelFinder module determined that a general time reversible (GTR) model with FreeRate site heterogeneity with 5 parameters provided the best fit to the data. The inferred tree is shown in Figure 7.3 along with isolate phylogroup and sequence types; in general, as expected, sequence types were largely monophyletic and phylogroups tended to cluster together.

7.3.3 Study isolates in a global context

The global collection of *E. coli* comprised 1273 samples, including the 473 from this study. 753/1253 (60%) were from Africa, 335/1253 (27%) from Asia and 167 (13%) from South America. The majority of samples, 1026/1253 (82%), were from stool, with 106/1253 (8%) truly invasive samples from blood or CSF and 63/1253 (5%) possibly invasive samples from urine, pus, or sputum. 65/1253 (5%) of samples were environmental, all from Thailand. 670/1253 (53%) of samples contained at least one ESBL-encoding gene. The majority of isolates with ESBL gene (622/670 [92%]) came from this study or the Thai ESBL study. Phylogroup A was the commonest phylogroup in the global collection (482/1273 [38%]), followed by B1 (333/1273 [26%]) and B2 (191/1273 [15%]); phylogroup C was uncommon in the global collection (74/1273 [6%]) but the majority of the phylogroup C samples came from this study (43/74 [58%]). All of these 43 phylogroup C isolates belonged to a single ST, ST410; this ST was not seen at all in the previous Malawian study of largely invasive

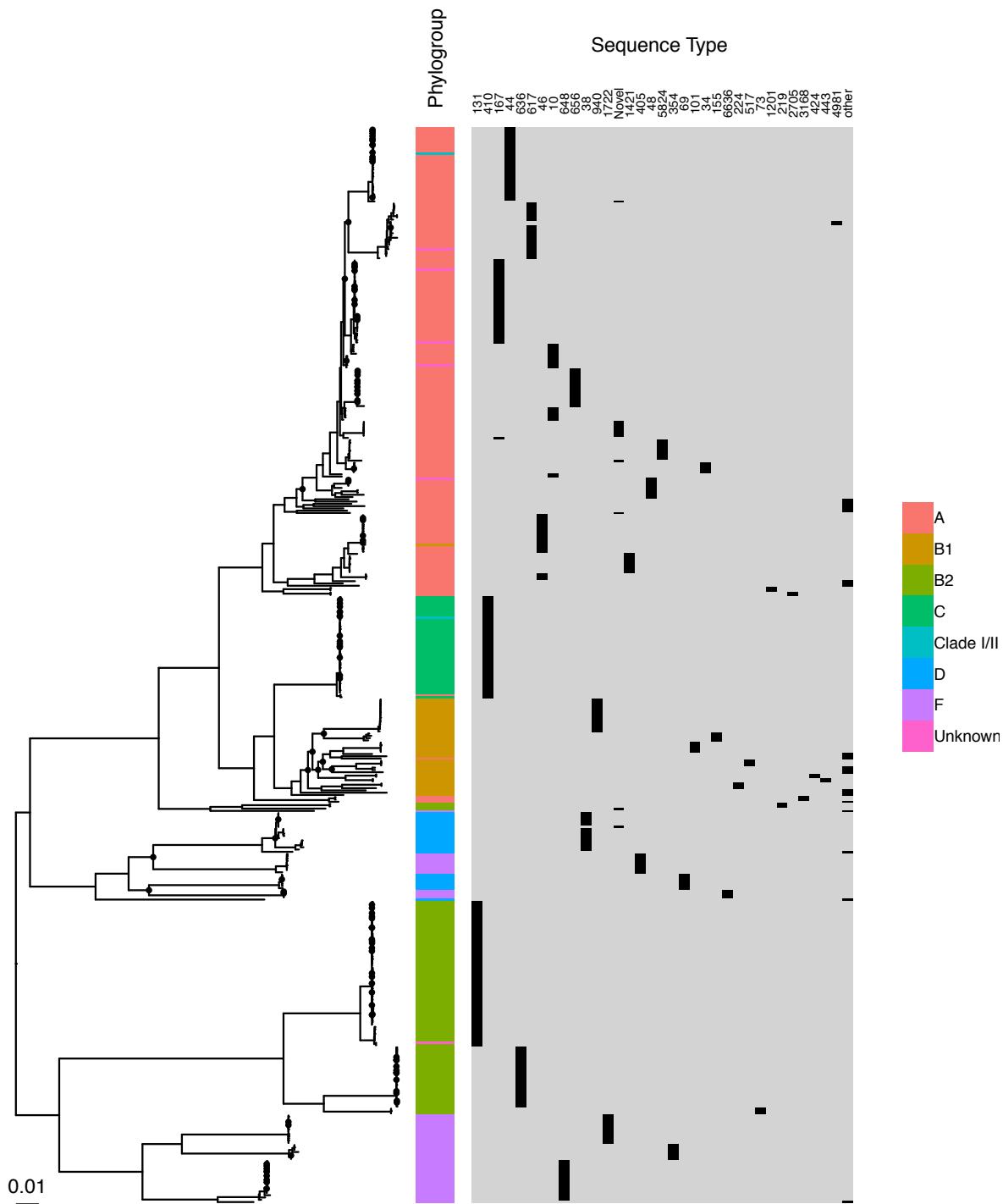


Figure 7.4: Maximum likelihood phylogenetic tree of included study *E. coli* isolates showing phylogroups and sequence types. Bootstrap support of less than 90% is indicated by a black circle at a given node. Scale bar indicates 0.01 SNPs/site.

isolates, despite being the second-commonest ST in this study, and was unusual in the global collection (11/800 [1%] ST410 in global collection vs 43/473 [9%] in this study). Similarly, the third-commonest ST in this study, ST167, was not seen at all in the global collection. However, ST131, the commonest ST in this study, was again the commonest ST in the global collection.

The Roary pan-genome pipeline identified 2872 core genes in a pan genome of 44840 genes; this large pan-genome is consistent with the open *E. coli* pan genome that will continue to increase in size as isolates are added. The core gene alignment contained 604817 bases with 77194 variable sites, which were used to infer the maximum likelihood phylogenetic tree, using same nucleotide substitution model as previously.

The inferred tree is shown in Figure 7.5). Isolates from this study are distributed throughout the tree, and there is widespread mixing of isolates from diverse geographic regions. Though invasive isolates are spread throughout the tree, there is a tendency for them to cluster together, particularly in phylogroup B2, a phylogroup with has a recognised association with ExPEC[22]. The Malawian ST410 and ST167 isolates clustered tightly together, but by comparison, ST131 isolates from this study were distributed among ST131 isolates from other studies, both in Malawi and elsewhere (Figure 7.6).

7.3.4 Antimicrobial resistance determinants

All identified AMR genes are shown in Figure 7.7A, alongside a summary of number of isolates with resistance mutations to given antimicrobial classes (Figure 7.7B) and the phenotypic resistance of the isolates for which phenotypic antimicrobial resistance testing was carried out (449/473 [95%]). The isolates contained a median (IQR) of 16 (12-17) resistance genes, and 100 different resistance alleles were identified in total. A description of resistance gene by class, along with a consideration of concordance (or otherwise) of phenotypic resistance and predicted resistance from genotype, are given in turn below.

7.3.4.1 β -lactam resistance

All isolates contained at least one gene that conferred resistance to third-generation cephalosporins, either an ESBL gene (n= 472) or a carbapenemase (n=1). The majority of ESBL-gene containing isolates contained only one ESBL gene (459/472 [97%]); fewer contained 2 (13/472 [3%]) and none contained more than 2. *bla_{CTX-M}* was the commonest ESBL gene, and over two thirds (319/473 [67%]) of isolates contained *bla_{CTXM-15}*. ESBL *bla_{SHV}* (26/473 [5%] of isolates) genes were also seen. ESBL *bla_{TEM}* (1/473 isolates) and

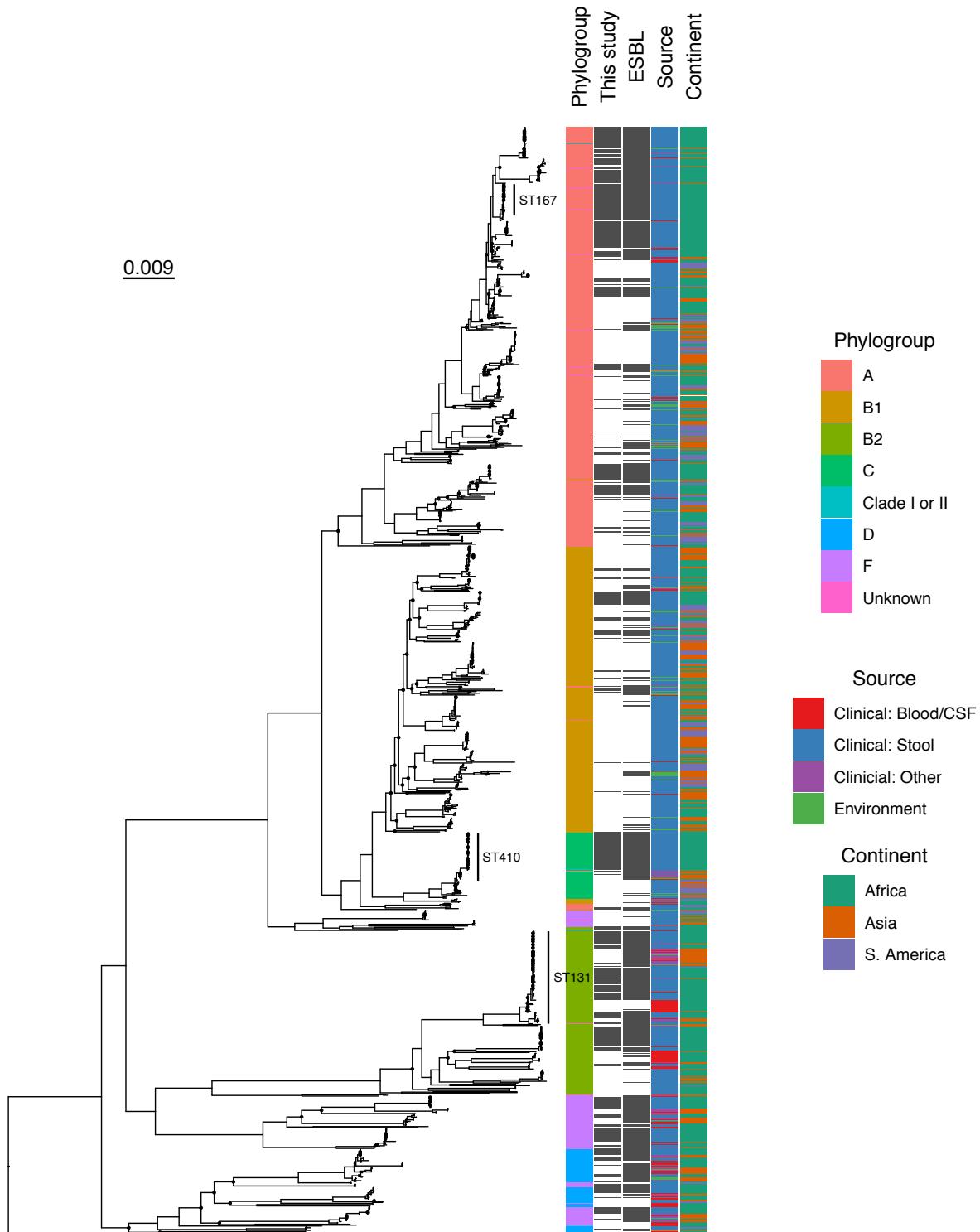


Figure 7.5: Midpoint rooted maximum likelihood phylogenetic tree of included study *E. coli* isolates along with global context isolates, showing phylogroups, source sample type and continent of isolation (coloured bars). Dark grey bars indicate isolates from this study or isolates with ESBL gene presence, as labelled (this study or ESBL, respectively). Three most frequently isolated STs in the current study (131, 410 and 167) labelled. Bootstrap support of less than 90% is indicated by a black circle at a given node. Scale bar indicates 0.009 SNPs/site.

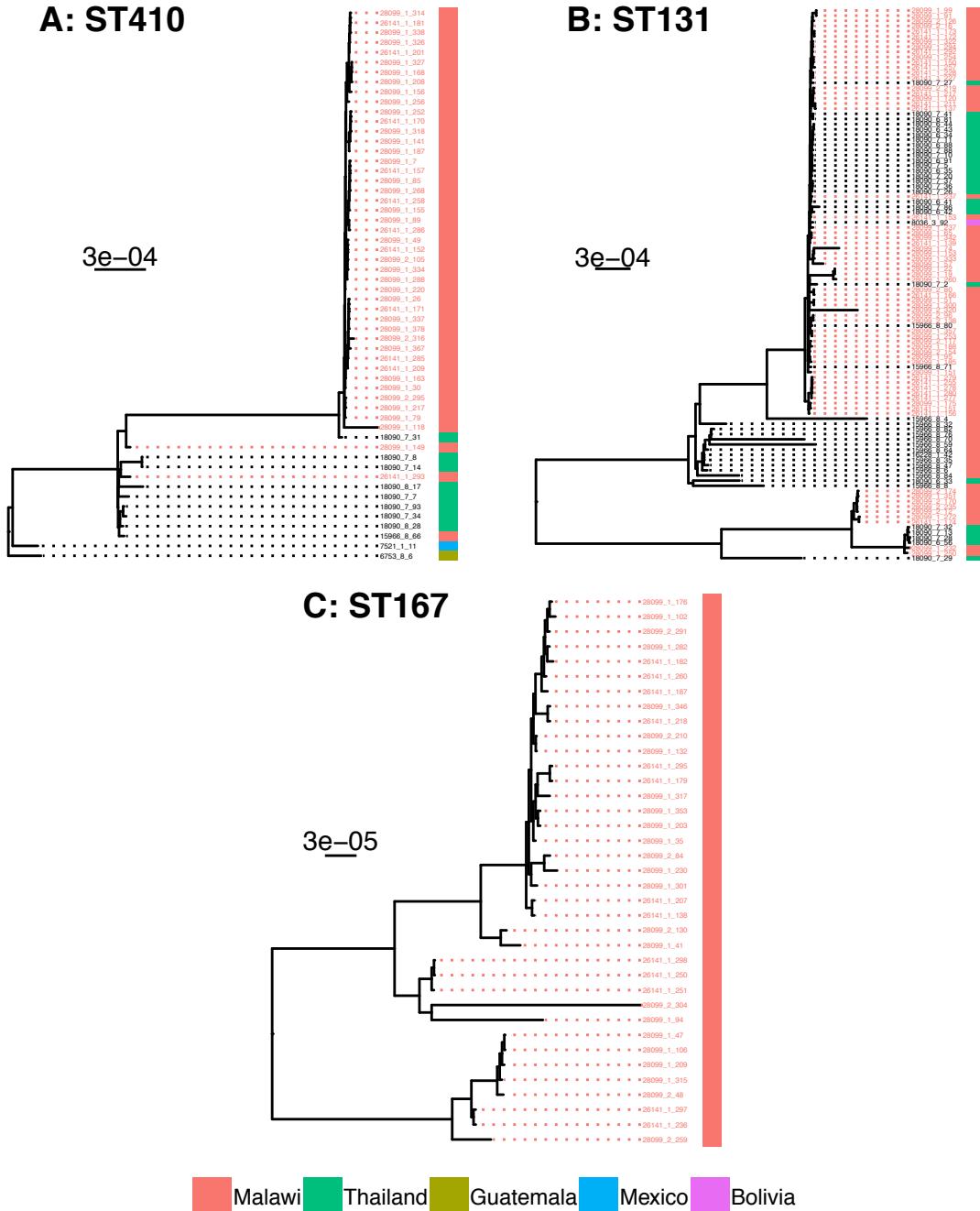


Figure 7.6: Subtree of ST410 (A, left) and ST131 (B, right), and ST167 (C, bottom) showing multiple introductions of ST131 into Malawi, in comparison to a single introduction clonal ST410 and ST167 clades. Colour of tree tip label indicates isolation from this study (red) or other studies (black), and coloured heatmap indicates country of isolation. Note that the scale bar in C is an order of magnitude different from A and B.

bla_{OXA} (1/473 isolates) were very unusual; however, narrow spectrum *bla_{TEM}* and *bla_{OXA}* β -lactamases were common: *bla_{OXA-1}* and *bla_{TEM-95}* were present in 186/473 [39%] and 289/473 [61%] of isolates respectively. Plasmid-mediated *bla_{ampC}* genes were identified in 45/473 (9%) of isolates, almost all (44/45) *bla_{CMY-42}*; this was unexpected as all of these isolates were confirmed to be ESBL-producers by combination disc testing. This testing uses cephalosporin-containing discs both with and without clavulanic acid, and confirms EBSL production by a difference in zone size between these discs, as ESBL enzymes are inactivated by clavulanic acid. However, the cephalosporins used in this test are likely to be hydrolysed by *ampC* enzymes, and if these isolates were producing such enzymes it could confer cephalosporin resistance regardless of the presence or absence of clavulanic acid. This was not the case for any of these isolates; none of them hydrolysed the cephalosporins used in the presence of clavulanic acid. It may be that the *bla_{CMY}* genes were not expressed.

The carbapenemase gene identified was a *bla_{NDM-5}*; the isolate harbouring this gene was recovered from the stool of a 67-year old man with no history of foreign travel nor hospitalisation. He had been admitted to the hospital with fever seven days previously and treated with seven days of intravenous ceftriaxone for sepsis, the source of which was not clear. He made an uneventful recovery, and no carbapenemase-containing isolate was recovered from his stool at any other time. The *bla_{NDM-5}* gene was carried on a partially assembled IncX3 plasmid. BLAST of this assembly against the NCBI database showed that this contig had 99% sequence identity with a previously sequenced pNDM-MGR194 46.2 kbp *bla_{NDM-5}* containing Inc-X3 plasmid found in India between 2011-13[23]. We fully assembled the plasmid by mapping reads back to pNDM-MGR194 with Burrows-Wheeler alignment and found it to be extremely similar, with only 13 SNPs compared to pNDM-MGR194.

7.3.4.2 Quinolone resistance

108/473 (23%) of isolates contained plasmid-mediated quinolone resistance PMQR genes, either *qnr* or *qep*. Nonsynonymous mutations were identified in at least one of the quinolone resistance-determining regions (QRDR) - *gyrA*, *gyrB*, *parC*, or *parE* - in 349/449 (78%) of isolates. The majority of mutations were well-described QRDR mutations (codon 83 and 87 in *gyrA*, codon 80 and 84 in *parC* and codon 458 in *parE*, Figure 7.8A). QRDR mutations tended to cluster together (Figure 7.8B) but alone they correlated poorly with phenotypic resistance. Of the 449 samples with available phenotypic sensitivity data, 294/449 (65%) were intermediate or resistant to ciprofloxacin, but 349/449 (78%) had a mutation in any codon in one of the four QRDR; presence of any QRDR mutation together with presence of PMQR had sensitivity of 95% (95% CI 93-98%) but specificity of 27% (95% CI 20-34%) for phenotypic quinolone resistance. Presence of mutations at all of codon 83 and 87 of

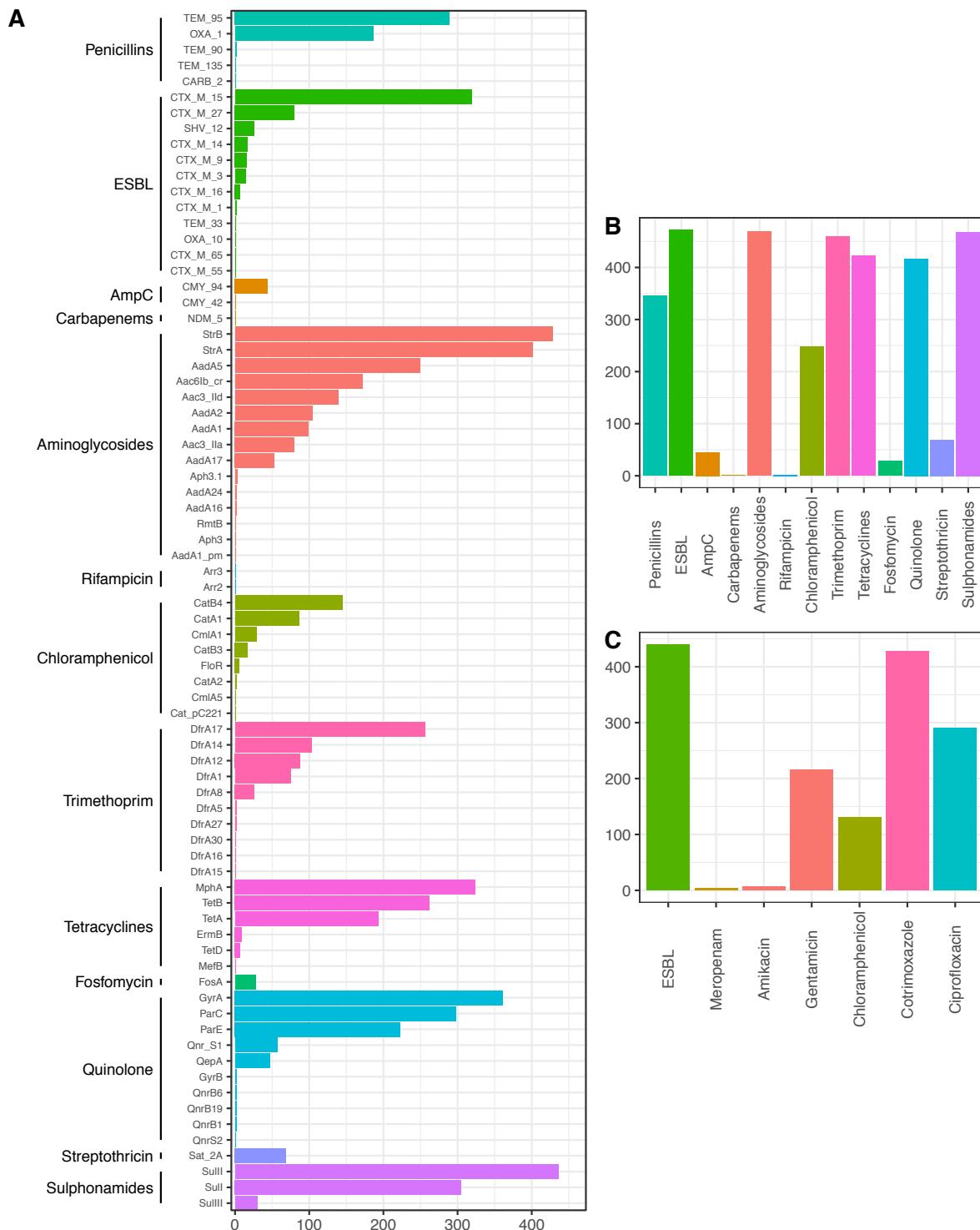


Figure 7.7: A: Frequency distribution of AMR genes identified in isolates. Class of antimicrobial to which gene confers resistance is shown. B: Number of isolates with any mutation to a given class. Any mutation that could possibly confer resistance to a given class is included, including any mutation in the QRDR for quinolones. C: Phenotypic resistance patterns for subset of samples in this analysis that also underwent phenotypic testing ($n = 449$)

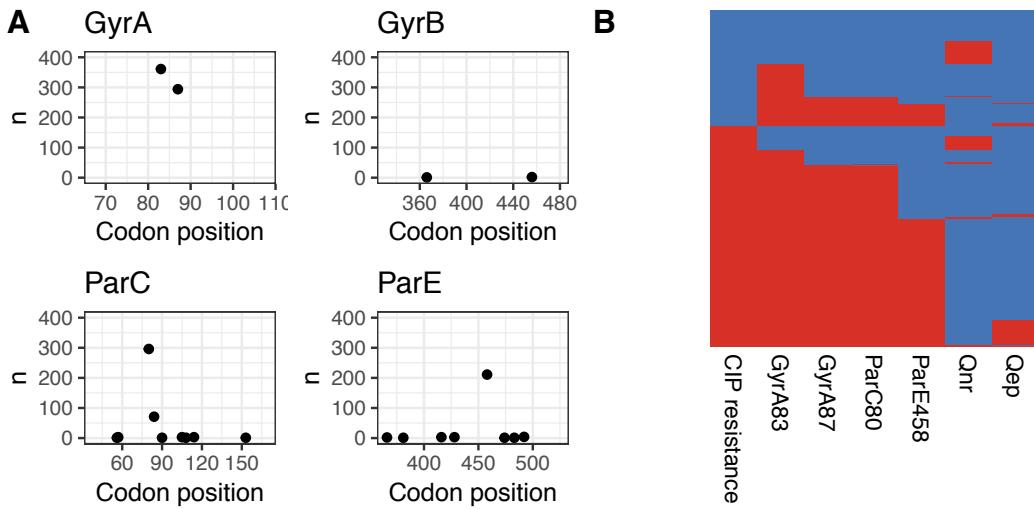


Figure 7.8: A: Mutation positions in quinolone resistance-determining regions, showing that most mutations are well-recognised (see text for details) B: Co-occurrence heatmap of QRDR mutations (*gyrA*, *parC*, or *parE*) plasmid-mediated quinolone resistance mutations (*qnr* or *qep*) and phenotypic resistance. Each row is one sample, red = presence, blue = absence.

gyrA and at codon 80 of *parC* has previously been shown to have the best predictive ability of phenotypic resistance[24], and this showed improved, but still poor, discrimination for phenotypic resistance with sensitivity 89% (95% 85-93%) and specificity 54% (95% CI 46-62%) in this dataset.

7.3.4.3 Aminoglycoside resistance

Aminoglycoside resistance genes were very common in the sequenced isolates, with 469/473 (99%) of isolates containing at least one aminoglycoside gene, and most containing multiple different genes: median number of aminoglycoside resistance genes per isolate was 4 (IQR 3-5). Despite streptomycin being absent from all Malawian treatment guidelines save for re treatment of tuberculosis, the streptomycin resistance genes *strA*, *strB* and *aadA* family of genes (also called *aad(3")*) were very commonly seen (Figure 7.9A). Genes that would be expected to confer gentamicin resistance - *aac(3)-IIa*, *aac(3)-IId* and *aac(6')-Ib-cr* were common, but genes that would be expected to confer amikacin resistance (*rmtB*) and kanamycin resistance (*aph(3')*) were unusual (Figure 7.9B)[25,26]

The predictive value of presence of *aac(3)-IIa*, *aac(3)-IId* or *aac(6')-Ib-cr* for phenotypic gentamicin resistance was moderate at best with sensitivity 77% (95% CI 71-83%) and specificity 73% (95% CI 67-79%). Of 6 phenotypically amikacin resistant or intermediate isolates, all had recognised streptomycin resistance determinants (*strA*, *strB* or *aadA*) but

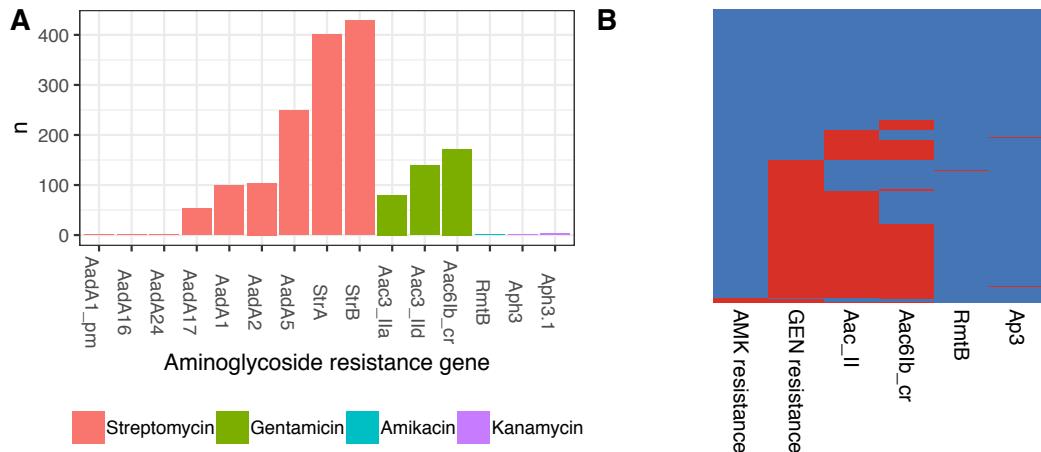


Figure 7.9: A: Aminoglycoside mutations and expected resistance to gentamicin, amikacin and kanamycin B: Heatmap showing phenotypic amikacin and gentamicin resistance and identified resistance genes that could be expected to confer resistance to these agents (see text for details). Aac_{II} in heatmap indicates presence of either *aac(3)-IIa* or *aac(3)-IIa* gene. Each row is one sample, red = presence, blue = absence.

4/6 had no other aminoglycoside determinant identified. Of the remaining two, one isolate contained *aac(6')-Ib-cr* and one both *aac(3)-IIa*, *aac(3)-IId*.

7.3.4.4 Chloramphenicol, co-trimoxazole, tetracycline and other resistance determinants

248/473 (52%) of isolates contained at least one chloramphenicol resistance gene (Figure 7.7), usually 1 (210/248 [85%]), less commonly 2 (37/248 [15%]) or 3 (1/248 [<1%]). *catB4* was the most commonly identified gene but once again phenotypic chloramphenicol resistance correlated poorly with presence of chloramphenicol resistance genes (7.10A)) with presence of any chloramphenicol resistance gene predicting phenotypic resistance with a sensitivity of 70% (95% 62-78%) and specificity of 55% (95% CI 49-60%).

Almost all isolates contained either a trimethoprim resistance (459/473 [97%]) or a sulphonamide resistance gene (468/473 [99%]); only 3/473 isolates did not contain either. Trimethoprim resistance genes were all of the *dfrA* family; *sulII* was the commonest sulphonamide resistance determinant (Figures 7.7 and 7.10B). Summary sensitivity of presence of any *dfrA* or *sul* gene as a predictor of phenotypic resistance was 100% [95% CI 99-100%] but (partially due to the rarity of co-trimoxazole sensitivity in this dataset) specificity was 13% [95% CI 2-40%].

Tetracycline resistance genes were also very common, identified in 422/473 (89%) of isolates,

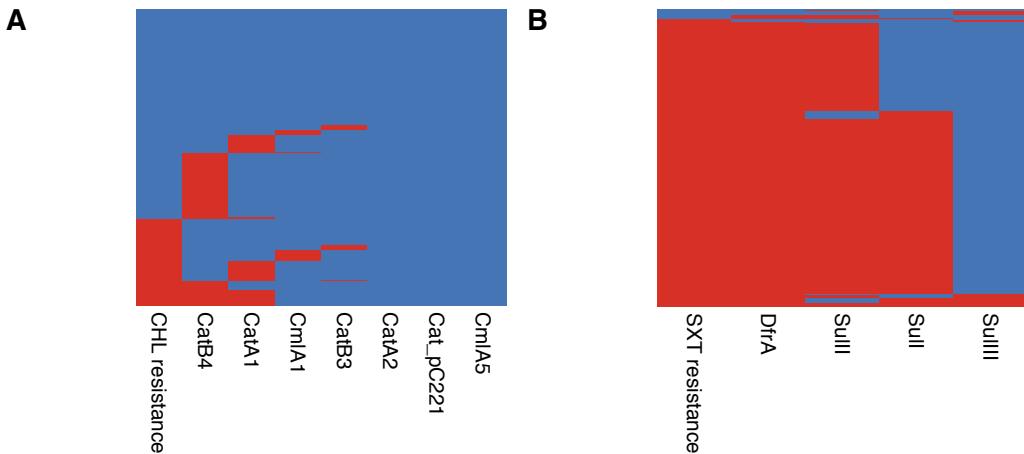


Figure 7.10: Heatmap showing phenotypic chloramphenicol (A) and cotrimoxazole (B) resistance and identified resistance genes that could be expected to confer resistance to these agents. Each row is one sample, red = presence, blue = absence

most commonly *mphA* (324/473 [68%] of isolates), followed by *tetB* (262/473 [55%] of isolates) and *tetA* (193/473 [41%] of isolates). No antimicrobial sensitivity testing was carried out for any agent of the tetracycline class. Resistance determinants for rifampicin (*arr2* and *arr3*) were rarely identified, in 2 isolates and the *sat2* gene, conferring resistance to streptothrinicin (a nucleoside antibiotic with no clinical compounds in use) was seen in 69/473 [15%] of isolates; the significance of this is unknown. Finally, the fosfomycin resistance determinant *fosA* was seen in 28/473 [6%] of isolates, despite this antimicrobial being unavailable in Malawi.

7.3.4.5 Clustering and lineage association of AMR determinants

Next, I explored associations of AMR determinants, both with each other in an attempt to identify putative clusters that could represent mobile genetic elements (MGE) that could be tracked within and between patients, and with lineages of the phylogeny. There was clear clustering of AMR genes beyond what would be expected by chance (Figures 7.11A and B), including clustering of the ESBL gene *bla_{CTXM-15}* with penicillinases *bla_{OXA-1}* and *bla_{TEM-95}*. Though some identified clusters correspond to known MGE (e.g. the *sulII-strA-strB* cluster[27]), there was a clear lineage association of certain gene combinations on mapping the presence or absence of AMR determinants back to the phylogeny (Figure ??C), meaning that these AMR-gene associations likely represent a combination of co-location on MGE and confounding by association with lineage.

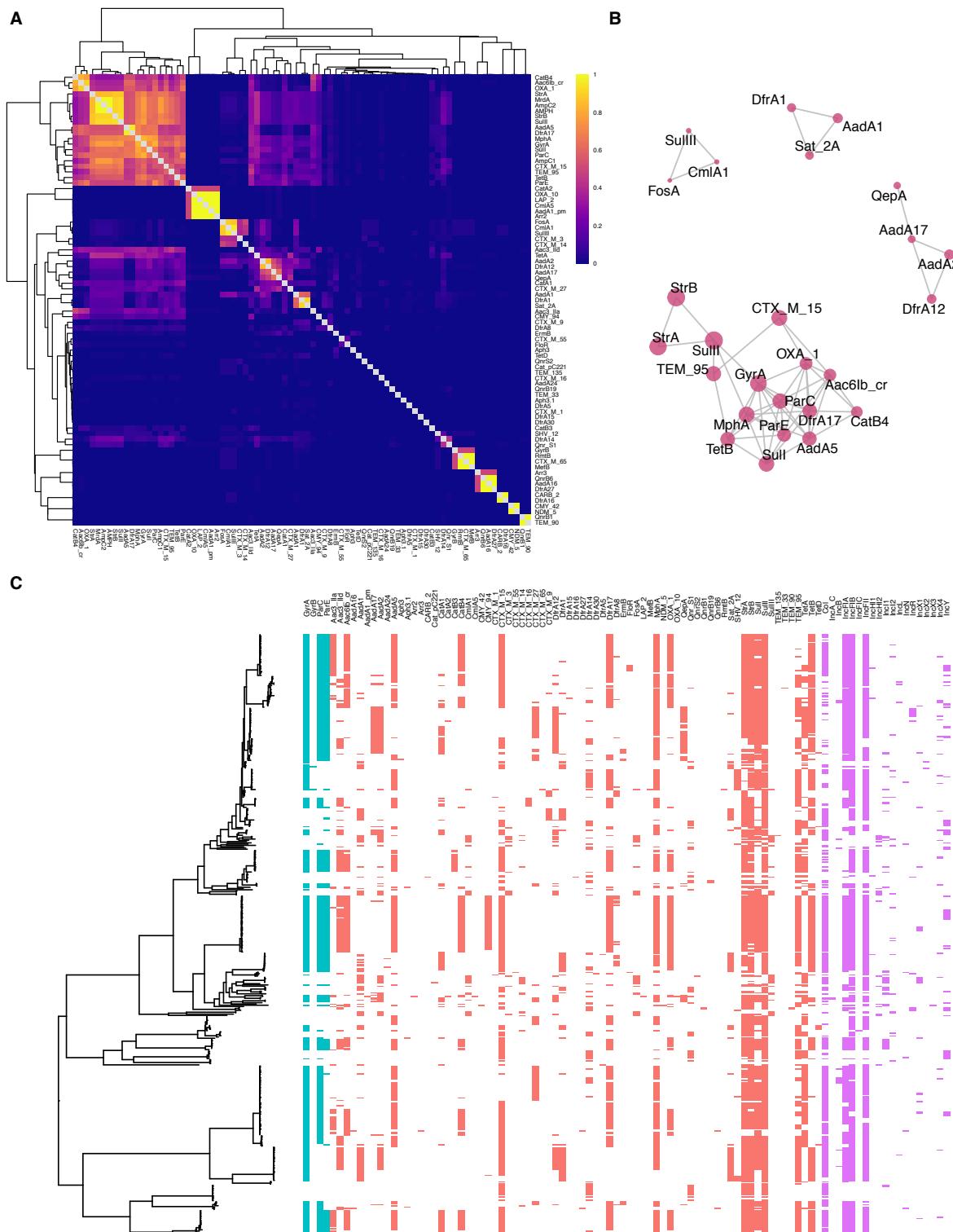


Figure 7.11: A: Row and column clustered heatmap of pairwise Jaccard index matrix, showing clustering of AMR genes. B: Networks of commonly (jaccard index > 0.5) and significantly ($p < 0.05$, Bonferroni corrected) co-occurring AMR genes. C: AMR genes mapped back to midpoint rooted maximum likelihood phylogenetic tree, showing lineage associations of genes.

7.3.5 Plasmid replicons

Presence or absence of the identified plasmid replicons is shown mapped to the phylogeny in Figure 7.11C. IncFIb was most commonly identified (399/473 [84%] of isolates), followed by IncFII (383/473 [81%] of isolates) and IncF1a (324/373 [68%] of isolates). Col plasmids were also frequently identified, in 308/473 [65%] of isolates. Once again, there seems to be some lineage associations of presence or absence of replicons.

7.3.6 Testing metadata associations: SNP distance, hierBAPS sequence clusters and ESBL-clusters

Finally, in order to test metadata associations of bacterial lineages or MGE, I used several techniques: considering core gene SNP distance between isolates to infer continuous carriage and/or transmission events, and clustering core gene pseudosequences and ESBL-containing contigs into mutually exclusive groups which can then be used to test associations. Below, I first describe the outcomes of the clustering algorithms used, before describing tests of association with metadata.

7.3.6.1 Hierarchical BAPS clustering of core gene pseudosequences

The hierarchical BAPS algorithm clustered the core gene alignments into 15 level one (top level) clusters, denoted sequence clusters A-O, and a total of 48 level two (lower level) clusters, denoted sequence clusters 1-48 that were almost exclusively monophyletic and often corresponded closely to the multilocus sequence types (STs, Figure 7.12A). Intracluster pairwise SNP distance varied (Figure 7.12B) but the clusters were often reasonably clonal: SC6, SC8 and SC23, for example (the three largest clusters) had median (IQR) intragroup pairwise SNP distance of 62 (34-97), 326 (18-378) and 18 (11-24) respectively.

7.3.6.2 ESBL-clusters

The 473 samples contained 486 ESBL genes (Figure 7.13A); 5 genes only occurred once in the collection and so no attempt was made to cluster them. Of the remaining 481 genes pairs, BLAST failed to identify the ESBL-gene containing contig in 2 samples (one in which ARIBA had identified *bla* – *CTXM* – 15 one *bla*_{CTXM-27}), but identified the remaining 479 ESBL genes on 478 contigs, with perfect agreement with ARIBA as to which AMR gene was present in which sample. Only one contig carried two ESBL genes: *bla*_{CTXM-3} and *bla*_{CTXM-15}; the remaining 477 contigs contained one. The *cd-hit* algorithm grouped the 477 unique contigs

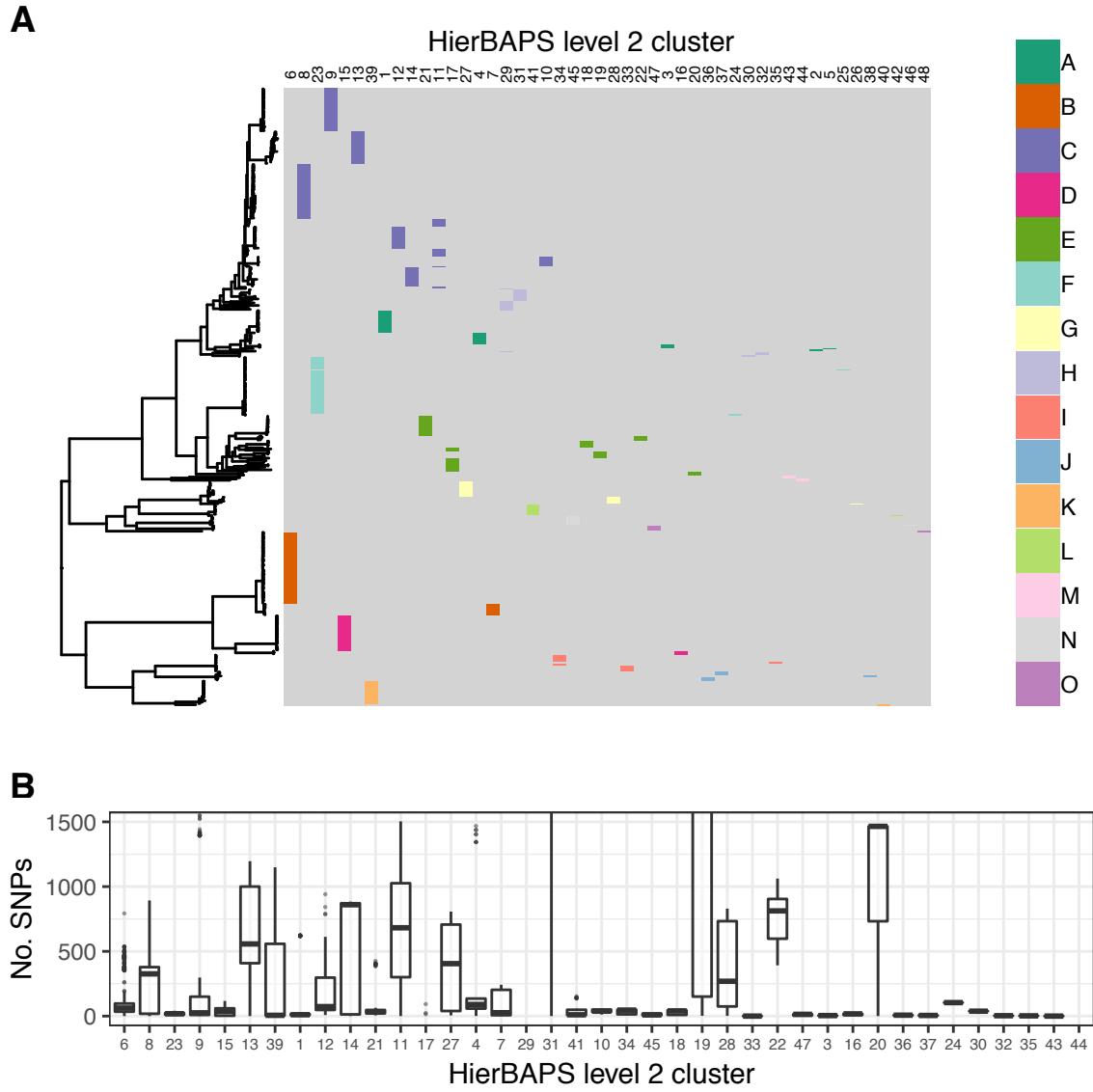


Figure 7.12: A: Core gene hierarchical BAPS clusters mapped back to phylogeny. Heatmap shows level 2 (lower level) with colour denoting level 1 (top level) cluster membership. B: Intracluster pairwise SNP distance for level 2 sequence clusters. Axis restricted to 0-1500 SNPs and as result SC17 (median 6881 SNPs), SC29 (median 2970 SNPs), SC31 (median 2970 SNPs) and SC44 (median 12322 SNPs) boxes are not shown. Boxplots show median and IQR, whiskers show 1.5 times IQR, and outliers are points falling beyond whiskers.

into 99 clusters (Figure 7.13B). In total, over 90% of the ESBL-genes (432/479 [90%]) were contained in the 52 largest contig clusters.

The *cd-hit* algorithm selects one member of a cluster (the longest) as the representative. The structure of these representative contigs was explored in an attempt to understand type of MGE they were likely to represent. The length of the representative clusters was very variable, ranging from 1.8kbp to 905.8kbp, with median (IQR) 46.1kbp (11.1-215.5kbp). The other cluster members were usually fragments of these representative contigs with varying sizes - a median (IQR) 60% (36-100%) of the representative contig length - but had high sequence identity, median (IQR) 100.0% (99.7-100.0%) (Figure 7.17 in the appendix to this chapter).

I then explored the insertion sequence (IS), AMR gene and plasmid replicon content of the representative contig for each cluster using BLAST against the SRST2, ISfinder and Plasmidfinder databases (Figures 7.18 and 7.19 in the appendix to this chapter). Every ESBL gene was closely associated with at least one IS, commonly ISEcp1, IS26 and IS903B. IS26 was frequently associated with an apparent 108bp fragment of a *catB4* chloramphenicol resistance determinant. Some ESBL-genes were associated with particular IS; *blaCTXM-15*, *blaCTXM-9* and *blaCTXM-1*, for example were very commonly associated with ISEcp1, whereas *blaSHV-12* was associated with IS26. ESBL genes were not infrequently associated with other resistance determinants, including commonly *blaCTXM-15* with *blaTEM-95*. Plasmid replicons were occasionally identified, including an IncFIB plasmid carrying *blaCTXM-15* and an IncQ1 plasmid carrying *blaCTXM-27*. It is clear that the same configuration of AMR genes and IS are seen across different contigs, despite a varying backbone, implying historical transposition events. Finally, to assess lineage associations of the identified ESBL-clusters, I mapped the clusters back to the tree, and found that there was a strong lineage association (Figure 7.13C).

7.3.6.3 Assessing for healthcare-associated lineages

Having clustered bacteria and MGE using *hierBAPS* and *cd-hit* respectively, I then mapped the location of sample collection back to the phylogeny and used the *hierBAPS* SCs to assess for healthcare associated lineages (Figure 7.14). In general, healthcare-associated isolates were distributed throughout the tree and across all SCs, rather than there being a clear hospital-associated lineage. The exception to this was SC23, corresponding to MLST 410, which was slightly more likely to be healthcare associated. When comparing the proportion of healthcare associated samples within each SC to the remained of samples, SC23 had a statistically significantly increased proportion of healthcare associated samples on Fisher's exact test ($p = 6.3 \times 10^{-4}$, threshold of significance following Bonferroni correction 1.0×10^{-3}), though it was by no means health-facility restricted: 50% (21/42) of SC23 samples were isolated in the community.

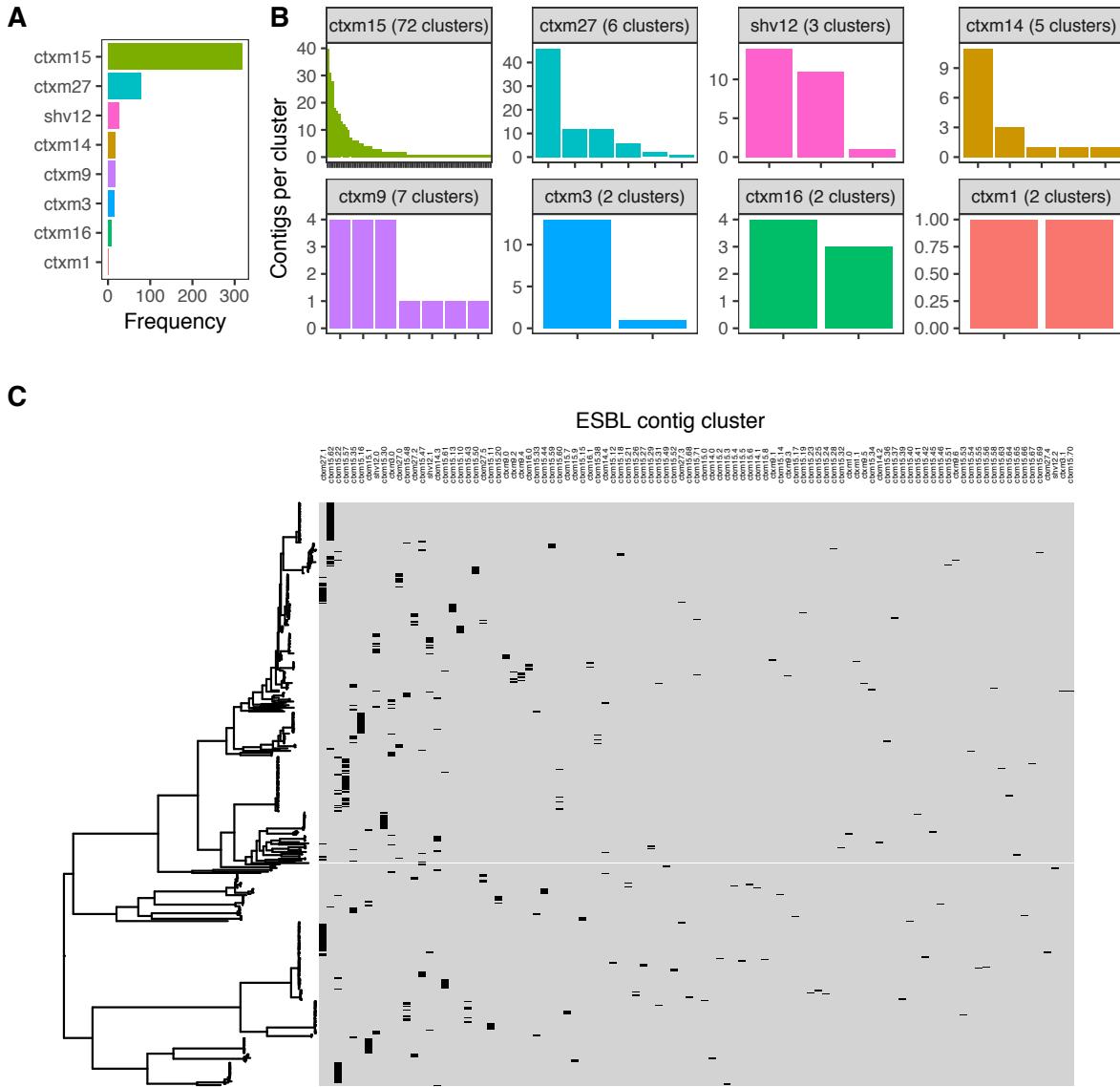


Figure 7.13: A: Frequency distribution of ESBL genes in included samples. B: Frequency distribution of samples per ESBL-cluster, stratified by gene. C: ESBL-cluster membership mapped back to phylogeny.

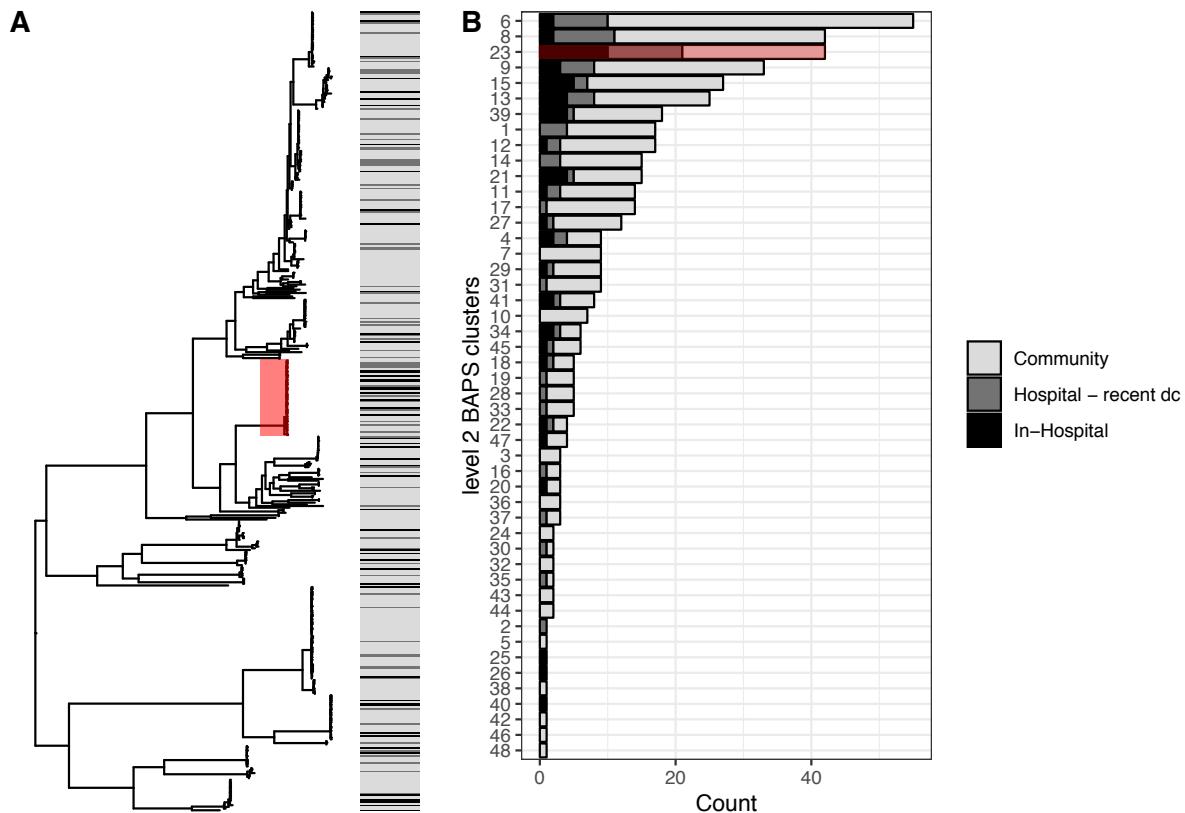


Figure 7.14: A: Location of sample isolation mapped back to phylogeny B: Distribution of location of sample isolation stratified by hierBAPS cluster. In each case, community isolates include those cultured from samples collected on the first day of hospital admission, in-hospital isolates are from patients who have been hospitalised > 24 hrs and recent discharge isolates are from patients who have been discharged from hospital within the last 2 weeks. Sequence cluster 23, highlighted in red, showed a statistically significant association with hospitalisation (see text).

7.3.6.4 Assessing for within-patient conservation of lineage or MGE

To answer the question as to what elements (bacteria or MGE) are conserved within individuals across time I first compared all-against-all pairwise SNP distance between and within patients; first as a scatter plot, and then, because of significant overplotting, as a density plot (Figure 7.15). This suggested that there are a cluster of points close to the origin in the within-patient plot that are not seen in the between-patient plot: before approximately 50 days, there are more similar within-patient isolates than seen in the between-patient isolates. Dichotomising time at 50 days (based on inspection of the density plots) and performing a Kruskal-Wallace test found a statistically significant difference between the before 50 day and after 50 day pairwise SNP distance distribution in the within patient stratum ($p = 0.008$) but not in the between-patient stratum ($p = 0.07$). After 50 days, the distribution of between- and within-patient SNP distances are similar ($p = 0.45$). However it is clear from the plots that even at small t and within-participant, there is significant diversity in the SNP distances, and that some isolates close together in time, within the same participant, are only distantly related.

Having confirmed that there is a signal for within-participant temporal conservation of ESBL-E, I then sought to determine if the sequence clusters and ESBL-clusters were similarly conserved over time, and if so, which was the more conserved. The proportion of pairwise within-patient samples that contained the same ESBL-cluster and sequence cluster were significantly greater than would be expected by chance when the time between the samples is less than 35 days for sequence cluster and 32 days for ESBL-cluster (Figure 7.16A). After this time, the lower confidence interval of the sequence cluster and ESBL-cluster curve crossed the proportion of samples that would be expected to be the same by chance, suggesting that, after 35 or 32 days, the chance of any two within-patient samples having the same sequence cluster or ESBL-cluster (respectively) is the same as if the two samples were randomly picked from the data set without regard to patient. The two curves have a very similar appearance; to address the question of which element is most conserved within an individual - sequence cluster, ESBL-cluster, or both - I performed an all-against-all pairwise comparison of which elements were conserved (Figure 7.16C), and found that only ESBL-cluster and sequence cluster together are conserved within patients at a significantly greater proportion than between patients ($p = 1.1 \times 10^{-12}$).

7.4 Discussion

In this chapter, I have presented the results of whole genome sequencing of 473 ESBL *E. coli* recovered from serial sampling of 230 Malawian adults from a combination of healthcare-associated and community settings, in an attempt to understand drivers of ESBL-E carriage. I

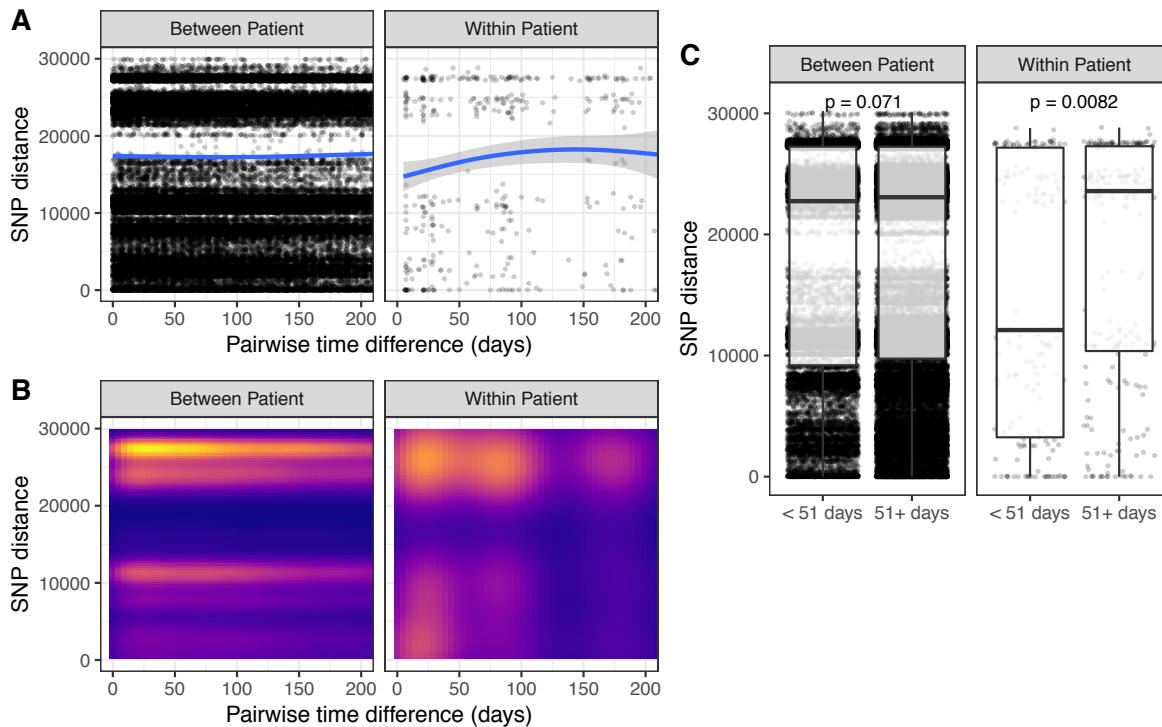


Figure 7.15: Within and between participant pairwise SNP distances. A: Scatterplot of pairwise SNP distances as a function of time with GAM model fitted curve. B: Pairwise SNP distance as function of time as a 2D density plot, showing cluster of isolates close to origin that are close together in time and SNP-distance. C: Pairwsise SNP distance distribution before and after 50 days, within and between patients, showing statistically significant descreas ein pairwise SNP distance within patients before 50 days. After 50 days, between- and within- patient distributuions are similar.

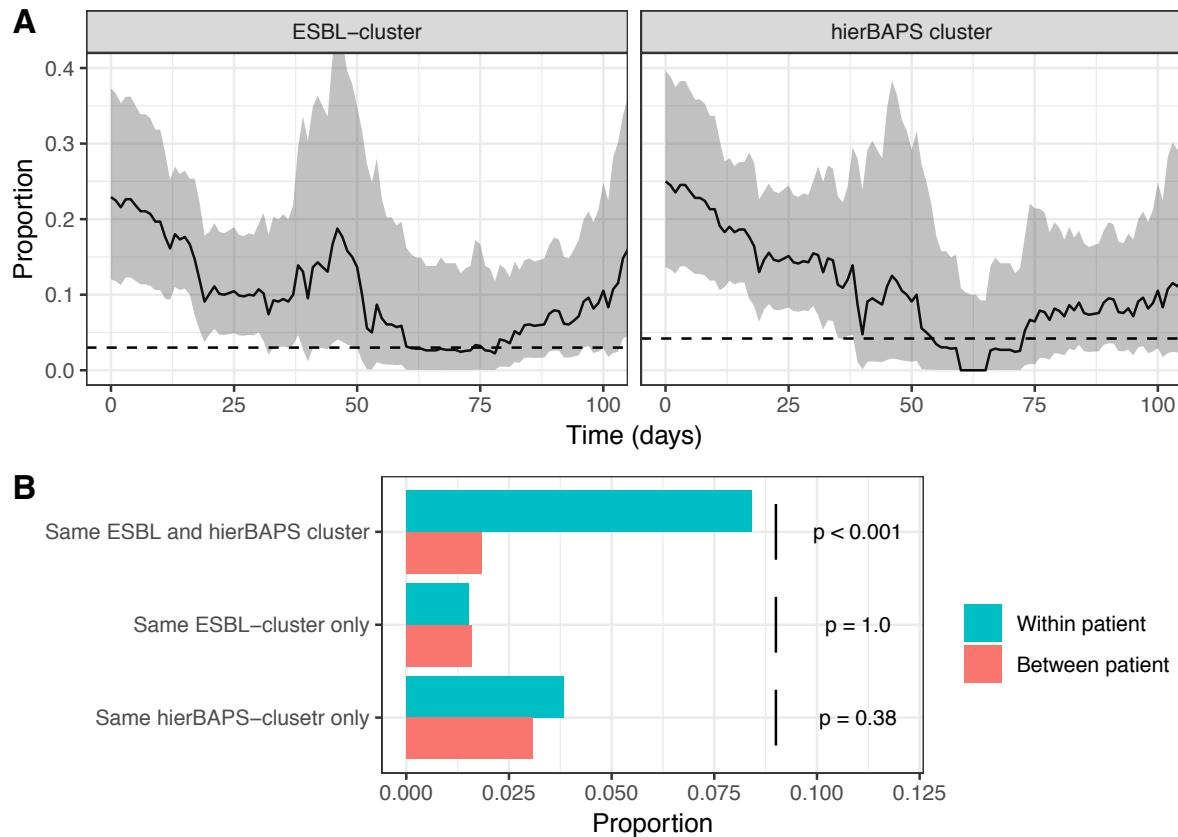


Figure 7.16: Probability of any two samples from within a given participant containing the same ESBL-cluster (A, left panel) or being a member of the same hierBAPS cluster (A, right panel). Time is windowed at ± 5 days around the time indicated on the x axis. Dotted line is the probability that two samples would belong to the same group by chance, constructed by randomly sampling 1000 sample pairs. B: proportion of samples that contain the same herBAPS cluster alone, or ESBL-cluster alone, or both, demonstrating that the ESBL cluster-hierBAPS cluster pairing is the most conserved of the three.

below discuss first the genomic landscape of the isolates in general terms, followed by insights that this analysis provides into the drivers of ESBL-E carriage.

7.4.1 Genomic landscape of ESBL *E. coli* in Malawi: global diversity and high-risk clones

The *E. coli* recovered from stool of the study participants in this study are diverse, encompassing the spectrum of diversity of the species with all major phylogroups and 56 STs represented. Phylogroup A was the commonest phylogroup seen, consistent with the traditional view of this phylogroup as associated with commensal strains[22]. When placed into a global phylogeny with context genomes from throughout the world, the Malawian isolates are largely distributed throughout the phylogeny: in a global context, Malawi is sampling the worldwide diversity of *E. coli*. The commonest ST in this study was the globally successful high risk clone ST131, mirroring the situation worldwide where this ST is thought to account for 40-80% of invasive ESBL *E. coli* infection[28,29].

There were, however, several areas of the global phylogeny where the Malawian isolates clustered tightly together, perhaps initially suggestive of Malawi-restricted clones; in considering the significance of this tree topology it is important to be cognisant of the biases inherent in the global *E. coli* collection, however. ESBL-producing *E. coli* are unusual in the ETEC[19] and GEMS[20] collections and all samples in these two studies were collected before 2011, though both of these collections are exclusively recovered from stool. In contrast, the clinical isolates from the Thai study[18] are all invasive, from a single centre, are selected on the basis of being ESBL-producers, and were isolated in 2014 or 2015. The isolates from the previous Malawian study were largely invasive[21], were selected for diversity in AMR profile, and were all isolated before 2014. There was no study that selectively cultured for ESBL producing *E. coli* in stool, as this study has done; in that, all of these studies are slightly sub-optimal for comparison. It may be that these biases contribute to apparent polyphyletic clustering of isolates from the current study in phylogroup A. It would be expected that ESBL producing phylogroup A *E. coli* would be underrepresented in the global collection compared to this study, as this phylogroup is associated with commensal (and hence stool) associated strains, and the two studies performing stool culture did not enrich for ESBL producers; the only study to do this collected invasive isolates, which may be expected to lie in phylogroup B2 over A.

Nevertheless, two of the three commonest STs identified in this study, ST410 and ST167, are unusual or absent in the global collection and could be considered to have a single introduction into Malawi in the context of the topology of the inferred phylogenetic tree. These could represent Malawi-restricted clades or, more likely given the diversity otherwise seen in the tree,

clades that are not represented in the global collection because of biases in sample selection. ST410 is recognised as an emerging high-risk clone which has been with isolated worldwide with some regularity since 2011 (including in Tanzania) and is associated with *bla_{CTXM-15}* and *bla_{NDM-5}*; coalescence analysis suggested a most recent common ancestor of ST410 of the early 1800s (similar to ST131[30]), and acquisition of *bla_{CTXM-15}* on a multireplicon IncFII-IncFIA-IncFIB plasmid in the late 1980s[31]. Similarly, ST167 has been recognised as commonly carrying ESBL genes and carbapenemases in Chinese invasive isolates[32] - indeed, it was the commonest *E. coli* ST in one longitudinal surveillance study of carbapenemases in 2012-16 in 25 Chinese provinces[33] - as well as being very prevalent among meat-associated *E. coli* in Germany between 2011-2013 in one study[34]. As such it, too, is also likely a successful global AMR-associated lineage.

It is therefore likely, especially given that these STs did not appear in the previous Malawian *E. coli* collection (in which all samples were collected prior to 2014), that they have recently arrived in Malawi. Alternatively, given that the previous Malawian collection was largely invasive, it could be that these STs do not often cause invasive disease, but the frequent identification of invasive ST410 and ST167 in the literature argues against this hypothesis, and it is more likely that they are recently arrived potential high-risk clones. If this is the case, then ST410 and 167 have become rapidly established in Blantyre over the course of only 2-3 years; in fact, this is exactly what was seen in longitudinal nationwide genomic surveillance of *E. coli* in the UK in 2003-04 when ST131 first arrived[35]. ST131 is now a well-established globally disseminated high-risk clone and the topography of the global tree suggests relatively unrestricted mixing between Malawian and global ST131. It may be that unbiased global sampling would reveal the same pattern for ST167 and ST410, and that they have also become globally prevalent; alternatively they may be truly geographically restricted either because they are in the process of spreading worldwide, or because some unknown factors prevent global spread. It is impossible to say from these data as they stand. Though some progress has been made in understanding the genomics of the emergence of ST131[30], the factors that contribute to its apparent fitness are unknown: it is impossible to predict, at present, from the genome of ST167 and ST410 whether they will repeat the course of ST131 to become truly globally dominant as a cause of human disease. Such an understanding of the determinants of fitness would be of great benefit in predicting and preventing global AMR spread.

7.4.2 Antimicrobial resistance determinants: domination of *bla_{CTXM-15}* and emergence of carbapenemases

The 473 isolates contained a diverse selection of antimicrobial resistance determinants, most with genotypic multiclass resistance. Genotypic and phenotypic co-trimoxazole resistance was

near universal, as might be expected from a setting where lifelong co-trimoxazole preventative therapy (CPT) is mandated by the Malawian HIV treatment guidelines for HIV all infected adults[36], and mediated by *dfrA* and *Sul* alleles. Determinants of quinolone resistance were commonly identified, more frequently mutations in the QRDR than plasmid-mediated; chloramphenicol resistance genes were also common, most often *catB*. However, both of these correlated only moderately with phenotypic resistance as determined by disc diffusion testing. This is unexpected, as presence of chloramphenicol resistance genes has been shown to correlate well with phenotypic resistance[37], including in a study of 94 Malawian invasive isolates[38], though in this Malawian collection (the same collection of isolates as were included in the global collection in the current study), *catB* genes were rarely seen. Similarly, the QRDR GyrA83–ParC80–GyrA87 combination was found to be strongly correlated with quinolone resistance in a study of 10099 *E. coli* genomes[24]. There are several potential explanations for these discrepancies; the first is that antimicrobial sensitivity testing (AST) has incorrectly classified sensitive and resistance isolates. The AST method used was disc diffusion; certainly this is less accurate than an MIC determination method such as E-tests or dilution methods, and a true comparison of genotypic to phenotypic resistance - not the aim of this study - should use one of these methods. It is also possible that there were technical problems with the AST procedure (e.g. an overly heavy inoculum) though every attempt was made to avoid this, including with internal QC, and the work was carried out in a laboratory which subscribes to the UK NEQUAS QC procedure.

Alternatively, a gene may be present but not expressed, or expressed at a low level. This can not be the case for the point mutations in the QRDR conferring quinolone resistance, but is possible for *catB4*, which was commonly present in phenotypically chloramphenicol sensitive isolates. Interestingly, truncated *catB4* elements (often in conjunction with an IS26 transposon) were almost universal in the isolates in this study: of 233 isolates in which ARIBA did not assemble a full *catB4* sequence, 226/233 (97%) contained a truncated *catB4* element. This configuration (*catB4* truncated by an IS26 element) has been described in Enterobacteriaceae[[39]; Sekizuka2018]. It could certainly be unrelated but its ubiquity in this study raises at least the possibility of misassembly and false-positive identification of *catB4* in some cases. Long read sequencing would allow description of the genetic environment of *catB4* in this collection and an understanding of its expression, which might allow better correlation with phenotype.

ESBL resistance in this collection is dominated by *blaCTXM* and *blaCTXM-15* in particular; this latter gene was carried by 319/473 [67%] of isolates. The only non-*blaCTXM* ESBL gene identified in any significant prevalence was *blaSHV-12*; ESBL *blaOXA* and *blaTEM* were very rare, though narrow-spectrum penicillinase alleles of this family were common. The dominance of *blaCTXM-15* is in keeping with the situation seen worldwide[40]. In this

collection, *blaCTXM-15* was spread throughout the phylogeny rather than associated with any particular clade. I identified one carbapenemase, *blaNDM-5*, carried on a globally successful IncX3 plasmid. To my knowledge, this is the first carbapenemase to be described in Malawi. Carbapenem antimicrobials were introduced to the Malawian essential medicines list in 2015 but are at best sporadically available, and only in tertiary centres, often for truncated courses. The emergence of carbapenemases with such minimal carbapenem use and so soon after introduction is troubling, and should prompt discussions regarding the best use of this precious antimicrobial class; certainly, given the high prevalence of ESBL production among invasive *Klebsiella pneumoniae* and *E. coli*[41], there is a case for expanded access but optimal antimicrobial stewardship strategies in this context are unknown.

7.4.3 Drivers of ESBL-E carriage: true acquisition versus enrichment

The diversity of healthcare associated isolates was largely contained within the diversity of community isolates, rather than apparent hospital acquisitions being restricted to a single clade or clone. The exception to this was SC23, which corresponded to ST410, and was more likely to be healthcare associated. This could be consistent with the hypothesis that it is a recently arrived high-risk clone, which may be, at least initially, hospital-associated. Even so, it is clearly not hospital restricted, with half of the ST410 isolates being isolated from the community. This result could be explained by two hypotheses: first, that these are true transmission events that are occurring within the hospital, and that the diversity of ESBL *E. coli* within the hospital is the same as the community; or, second, that these “hospital acquisitions” are actually minority variant *E. coli* present in the microbiota (and therefore acquired in the community) at a low abundance and hence not detected by culture, and enriched for with antimicrobial exposure in hospital. This latter hypothesis is perhaps supported by the fact that the antimicrobial-exposed group of patients in this study have an early and rapid increase in ESBL-E carriage prevalence that is not seen in the antimicrobial-unexposed group. Distinguishing between these two hypotheses is important as they would each require a different intervention: hospital infection control in the former case, or strategies to protect the microbiota from the deleterious effect of broad spectrum antimicrobials (such as pre- or probiotics, or oral β -lactamases) in the latter.

By forming sequence clusters and ESBL-clusters, I was able to demonstrate that both bacteria and MGE are conserved together, within-patient, over time. Some previous longitudinal studies of ESBL-E found that *E. coli* STs tended to vary over time but that in many cases ESBL gene and plasmid replicon remained the same, which could be due to a conserved MGE transferring between bacteria[42]. Given my findings, this is unlikely to be the case. Though not directly addressed in this study it is possible to speculate therefore that the unit of

transmission of ESBL between patients is likely to be the bacterium rather than, for example, horizontal gene transfer of ESBL genes on plasmids or transposons. The within-patient correlation of SC and ESBL-cluster lasts only for 32-35 days; two samples from a single patient more than 35 days apart are as likely to contain the same SC/ESBL cluster as two samples from two different patients. This implies either an exogenous re-exposure or some other endogenous mechanism whereby the dominant ESBL strain is replaced by a minority variant from within the microbiota. Of note, the timescale of SC replacement - occurring after 35 days - is consistent with the mean time taken to revert to the ESBL-negative state from ESBL-positive from the longitudinal Markov models (26 [95% CI 12-58] days), perhaps lending support to the re-exposure hypothesis. It is important to note also that even at a maximum (i.e. with samples that are days apart), only around 20% of within-patient samples contain the same SC/ESBL-cluster. This is many times more than would be expected by chance, but still implies that at any time point there is significant diversity of ESBL *E. coli* strains, that have been missed by only taking forward one colony pick for sequencing. This is consistent with studies that have performed multiple colony picks on stool samples enriched for ESBL-E, and have found widespread diversity of STs and ESBL genes[43].

7.4.4 Study limitations

There are several limitations to the analysis carried out in this chapter. The most serious is due to the fact that only one colony pick from the ESBL selective media was taken forward for sequencing. This was inevitable because of resource considerations, but it is clear that there is significant within-patient unsampled diversity as a result. In effect, we have randomly sampled one strain from all available strains at any give time point. This is likely to result in an underestimation of the extent to which strains persist within the individual over time, as strains that are present (but not sampled) are classed as absent. I have focused on *E. coli*; in fact, these were often isolated in conjunction with other bacteria, most notably *Klebsiella pneumoniae*, which were not included in this analysis, so the problem of unsampled diversity is in fact even more severe. The diversity of ESBL genes in Blantyre is likely therefore to be greater than I describe.

I selected a global collection of *E. coli* based on what was available but, as described above and in common with many analyses of this type, this is a biased collection. This must be borne in mind when interpreting the global phylogeny. There are inherent limitations in the short-read Illumina sequencing that was carried out: assembly of areas with multiple nucleotide repeats (as found in plasmids and transposable elements in particular) is difficult or impossible, making it impossible to fully characterise the MGE in this dataset upon which the AMR genes are carried. I have attempted to address this difficulty by defining ESBL-clusters

as a proxy for MGE, but this is by its nature flawed. Some of the assembled contigs are short and likely represent transposons; the same transposons have likely inserted into multiple plasmids in the past and as such, these short contigs may cluster with other sequences that would be seen to be very different, were a full assembly available. In addition, the biological significance of these ESBL-clusters is not clear. It is not possible to say with certainty what they represent (e.g. plasmid) as they are only fragments. Nevertheless, the fact that I have seen within-patient associations of the ESBL-clusters lends some support to their use, as erroneous clustering would be expected to bias any associations towards to null.

7.4.5 Conclusions and further work

In conclusion, I have shown that the *E. coli* population in this study is diverse, representing global *E. coli* diversity; on a smaller scale, the diversity of healthcare-associated isolates is represented in the community. These facts together suggest widespread mixing of strains at multiple spatial levels. It is possible that apparent constant colonisation actually represents frequent re-exposure on the timescale of around 35 days, and that apparent hospital acquisitions, are, in fact, unmasking events due to enrichment of the microbiota for ESBL-E by antimicrobial exposure.

Many questions remain unanswered and further work is necessary. Shotgun metagenomic sequencing of stool would allow testing of the competing acquisition and unmasking hypotheses of rapid increase in ESBL-E prevalence by defining the microbiota and total AMR gene content pre-, during and post- antimicrobial exposure, as well as providing an opportunity to explore the role of the microbiota to colonisation resistance to ESBL-E. Long read sequencing would allow a proper characterisation of the MGE that carry ESBL genes in the Malawian context, giving the resolution necessary to truly track MGE within and between patients and strains, as well as to address questions of expression of genes such as *catB4* that seem to correlate poorly with phenotype, by examining e.g. promoter regions. Short-read sequencing of the *Klebsiella pneumoniae* isolates from this study would allow a comparison between the mechanisms of AMR and MGE prevalent in this species as compared to *E. coli*, and assess the extent to which horizontal gene transfer between the two is driving ESBL spread in Blantyre. Collating publicly available ST167 and ST410 genomes would allow the construction of a phylogeny that would allow insight into the epidemiology of these putative recently arrived high risk clones. Finally, incorporating the resolution afforded by sequencing into the longitudinal Markov models may provide new insights in to the dynamics of ESBL-E carriage. This is the subject of the next chapter.

7.5 Appendix

Table 7.1: Details of samples included in global phylogeny

Study	Sample ID	Accession No.	Source	Country
This study	26141_1_134		Stool	Malawi
This study	26141_1_135		Stool	Malawi
This study	26141_1_136		Stool	Malawi
This study	26141_1_137		Stool	Malawi
This study	26141_1_138		Stool	Malawi
This study	26141_1_139		Stool	Malawi
This study	26141_1_140		Stool	Malawi
This study	26141_1_141		Stool	Malawi
This study	26141_1_142		Stool	Malawi
This study	26141_1_143		Stool	Malawi
This study	26141_1_144		Stool	Malawi
This study	26141_1_145		Stool	Malawi
This study	26141_1_146		Stool	Malawi
This study	26141_1_147		Stool	Malawi
This study	26141_1_148		Stool	Malawi
This study	26141_1_149		Stool	Malawi
This study	26141_1_150		Stool	Malawi
This study	26141_1_151		Stool	Malawi
This study	26141_1_152		Stool	Malawi
This study	26141_1_153		Stool	Malawi
This study	26141_1_154		Stool	Malawi
This study	26141_1_155		Stool	Malawi
This study	26141_1_156		Stool	Malawi
This study	26141_1_157		Stool	Malawi
This study	26141_1_158		Stool	Malawi
This study	26141_1_159		Stool	Malawi
This study	26141_1_160		Stool	Malawi
This study	26141_1_161		Stool	Malawi
This study	26141_1_162		Stool	Malawi
This study	26141_1_164		Stool	Malawi

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
This study	26141_1_165		Stool	Malawi
This study	26141_1_166		Stool	Malawi
This study	26141_1_168		Stool	Malawi
This study	26141_1_169		Stool	Malawi
This study	26141_1_170		Stool	Malawi
This study	26141_1_171		Stool	Malawi
This study	26141_1_172		Stool	Malawi
This study	26141_1_173		Stool	Malawi
This study	26141_1_174		Stool	Malawi
This study	26141_1_175		Stool	Malawi
This study	26141_1_176		Stool	Malawi
This study	26141_1_177		Stool	Malawi
This study	26141_1_178		Stool	Malawi
This study	26141_1_179		Stool	Malawi
This study	26141_1_180		Stool	Malawi
This study	26141_1_181		Stool	Malawi
This study	26141_1_182		Stool	Malawi
This study	26141_1_183		Stool	Malawi
This study	26141_1_184		Stool	Malawi
This study	26141_1_186		Stool	Malawi
This study	26141_1_187		Stool	Malawi
This study	26141_1_189		Stool	Malawi
This study	26141_1_190		Stool	Malawi
This study	26141_1_191		Stool	Malawi
This study	26141_1_192		Stool	Malawi
This study	26141_1_193		Stool	Malawi
This study	26141_1_194		Stool	Malawi
This study	26141_1_195		Stool	Malawi
This study	26141_1_196		Stool	Malawi
This study	26141_1_197		Stool	Malawi
This study	26141_1_198		Stool	Malawi

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
This study	26141_1_199		Stool	Malawi
This study	26141_1_200		Stool	Malawi
This study	26141_1_201		Stool	Malawi
This study	26141_1_202		Stool	Malawi
This study	26141_1_203		Stool	Malawi
This study	26141_1_204		Stool	Malawi
This study	26141_1_205		Stool	Malawi
This study	26141_1_206		Stool	Malawi
This study	26141_1_207		Stool	Malawi
This study	26141_1_208		Stool	Malawi
This study	26141_1_209		Stool	Malawi
This study	26141_1_210		Stool	Malawi
This study	26141_1_211		Stool	Malawi
This study	26141_1_212		Stool	Malawi
This study	26141_1_213		Stool	Malawi
This study	26141_1_214		Stool	Malawi
This study	26141_1_215		Stool	Malawi
This study	26141_1_217		Stool	Malawi
This study	26141_1_218		Stool	Malawi
This study	26141_1_219		Stool	Malawi
This study	26141_1_220		Stool	Malawi
This study	26141_1_221		Stool	Malawi
This study	26141_1_222		Stool	Malawi
This study	26141_1_223		Stool	Malawi
This study	26141_1_224		Stool	Malawi
This study	26141_1_225		Stool	Malawi
This study	26141_1_226		Stool	Malawi
This study	26141_1_227		Stool	Malawi
This study	26141_1_228		Stool	Malawi
This study	26141_1_229		Stool	Malawi
This study	26141_1_230		Stool	Malawi
This study	26141_1_232		Stool	Malawi

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
This study	26141_1_236		Stool	Malawi
This study	26141_1_237		Stool	Malawi
This study	26141_1_239		Stool	Malawi
This study	26141_1_240		Stool	Malawi
This study	26141_1_241		Stool	Malawi
This study	26141_1_242		Stool	Malawi
This study	26141_1_243		Stool	Malawi
This study	26141_1_244		Stool	Malawi
This study	26141_1_246		Stool	Malawi
This study	26141_1_247		Stool	Malawi
This study	26141_1_248		Stool	Malawi
This study	26141_1_250		Stool	Malawi
This study	26141_1_251		Stool	Malawi
This study	26141_1_252		Stool	Malawi
This study	26141_1_253		Stool	Malawi
This study	26141_1_254		Stool	Malawi
This study	26141_1_255		Stool	Malawi
This study	26141_1_256		Stool	Malawi
This study	26141_1_257		Stool	Malawi
This study	26141_1_258		Stool	Malawi
This study	26141_1_259		Stool	Malawi
This study	26141_1_260		Stool	Malawi
This study	26141_1_261		Stool	Malawi
This study	26141_1_262		Stool	Malawi
This study	26141_1_263		Stool	Malawi
This study	26141_1_265		Stool	Malawi
This study	26141_1_266		Stool	Malawi
This study	26141_1_267		Stool	Malawi
This study	26141_1_268		Stool	Malawi
This study	26141_1_270		Stool	Malawi
This study	26141_1_271		Stool	Malawi
This study	26141_1_272		Stool	Malawi

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
This study	26141_1_273		Stool	Malawi
This study	26141_1_274		Stool	Malawi
This study	26141_1_275		Stool	Malawi
This study	26141_1_276		Stool	Malawi
This study	26141_1_277		Stool	Malawi
This study	26141_1_278		Stool	Malawi
This study	26141_1_279		Stool	Malawi
This study	26141_1_280		Stool	Malawi
This study	26141_1_282		Stool	Malawi
This study	26141_1_283		Stool	Malawi
This study	26141_1_284	ERR3168700	Stool	Malawi
This study	26141_1_285		Stool	Malawi
This study	26141_1_286		Stool	Malawi
This study	26141_1_287		Stool	Malawi
This study	26141_1_288		Stool	Malawi
This study	26141_1_289		Stool	Malawi
This study	26141_1_290		Stool	Malawi
This study	26141_1_291		Stool	Malawi
This study	26141_1_292		Stool	Malawi
This study	26141_1_293		Stool	Malawi
This study	26141_1_295		Stool	Malawi
This study	26141_1_296		Stool	Malawi
This study	26141_1_297		Stool	Malawi
This study	26141_1_298		Stool	Malawi
This study	26141_1_299		Stool	Malawi
This study	28099_1_1		Stool	Malawi
This study	28099_1_10		Stool	Malawi
This study	28099_1_100		Stool	Malawi
This study	28099_1_102		Stool	Malawi
This study	28099_1_103		Stool	Malawi
This study	28099_1_104		Stool	Malawi

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
This study	28099_1_106		Stool	Malawi
This study	28099_1_107		Stool	Malawi
This study	28099_1_11		Stool	Malawi
This study	28099_1_110		Stool	Malawi
This study	28099_1_111		Stool	Malawi
This study	28099_1_112		Stool	Malawi
This study	28099_1_114		Stool	Malawi
This study	28099_1_115		Stool	Malawi
This study	28099_1_116		Stool	Malawi
This study	28099_1_118		Stool	Malawi
This study	28099_1_119		Stool	Malawi
This study	28099_1_120		Stool	Malawi
This study	28099_1_123		Stool	Malawi
This study	28099_1_125		Stool	Malawi
This study	28099_1_127		Stool	Malawi
This study	28099_1_128		Stool	Malawi
This study	28099_1_129		Stool	Malawi
This study	28099_1_131		Stool	Malawi
This study	28099_1_132		Stool	Malawi
This study	28099_1_133		Stool	Malawi
This study	28099_1_135		Stool	Malawi
This study	28099_1_136		Stool	Malawi
This study	28099_1_137		Stool	Malawi
This study	28099_1_139		Stool	Malawi
This study	28099_1_14		Stool	Malawi
This study	28099_1_141		Stool	Malawi
This study	28099_1_143		Stool	Malawi
This study	28099_1_144		Stool	Malawi
This study	28099_1_145		Stool	Malawi
This study	28099_1_148		Stool	Malawi
This study	28099_1_149		Stool	Malawi
This study	28099_1_151		Stool	Malawi

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
This study	28099_1_152		Stool	Malawi
This study	28099_1_153		Stool	Malawi
This study	28099_1_155		Stool	Malawi
This study	28099_1_156		Stool	Malawi
This study	28099_1_157		Stool	Malawi
This study	28099_1_159		Stool	Malawi
This study	28099_1_160		Stool	Malawi
This study	28099_1_161		Stool	Malawi
This study	28099_1_163		Stool	Malawi
This study	28099_1_165		Stool	Malawi
This study	28099_1_167		Stool	Malawi
This study	28099_1_168		Stool	Malawi
This study	28099_1_169		Stool	Malawi
This study	28099_1_171		Stool	Malawi
This study	28099_1_172		Stool	Malawi
This study	28099_1_173		Stool	Malawi
This study	28099_1_175		Stool	Malawi
This study	28099_1_176		Stool	Malawi
This study	28099_1_177		Stool	Malawi
This study	28099_1_179		Stool	Malawi
This study	28099_1_18		Stool	Malawi
This study	28099_1_180		Stool	Malawi
This study	28099_1_181		Stool	Malawi
This study	28099_1_185		Stool	Malawi
This study	28099_1_187		Stool	Malawi
This study	28099_1_188		Stool	Malawi
This study	28099_1_189		Stool	Malawi
This study	28099_1_19		Stool	Malawi
This study	28099_1_191		Stool	Malawi
This study	28099_1_192		Stool	Malawi
This study	28099_1_193		Stool	Malawi
This study	28099_1_195		Stool	Malawi

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
This study	28099_1_196		Stool	Malawi
This study	28099_1_199		Stool	Malawi
This study	28099_1_2		Stool	Malawi
This study	28099_1_200		Stool	Malawi
This study	28099_1_203		Stool	Malawi
This study	28099_1_204		Stool	Malawi
This study	28099_1_205		Stool	Malawi
This study	28099_1_207		Stool	Malawi
This study	28099_1_208		Stool	Malawi
This study	28099_1_209		Stool	Malawi
This study	28099_1_211		Stool	Malawi
This study	28099_1_212		Stool	Malawi
This study	28099_1_213		Stool	Malawi
This study	28099_1_214		Stool	Malawi
This study	28099_1_216		Stool	Malawi
This study	28099_1_217		Stool	Malawi
This study	28099_1_218		Stool	Malawi
This study	28099_1_22		Stool	Malawi
This study	28099_1_220		Stool	Malawi
This study	28099_1_221		Stool	Malawi
This study	28099_1_222		Stool	Malawi
This study	28099_1_224		Stool	Malawi
This study	28099_1_225		Stool	Malawi
This study	28099_1_226		Stool	Malawi
This study	28099_1_228		Stool	Malawi
This study	28099_1_229		Stool	Malawi
This study	28099_1_23		Stool	Malawi
This study	28099_1_230		Stool	Malawi
This study	28099_1_232		Stool	Malawi
This study	28099_1_233		Stool	Malawi
This study	28099_1_234		Stool	Malawi

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
This study	28099_1_236		Stool	Malawi
This study	28099_1_237		Stool	Malawi
This study	28099_1_238		Stool	Malawi
This study	28099_1_240		Stool	Malawi
This study	28099_1_241		Stool	Malawi
This study	28099_1_242		Stool	Malawi
This study	28099_1_244		Stool	Malawi
This study	28099_1_245		Stool	Malawi
This study	28099_1_246		Stool	Malawi
This study	28099_1_248		Stool	Malawi
This study	28099_1_249		Stool	Malawi
This study	28099_1_250		Stool	Malawi
This study	28099_1_252		Stool	Malawi
This study	28099_1_253		Stool	Malawi
This study	28099_1_254		Stool	Malawi
This study	28099_1_256		Stool	Malawi
This study	28099_1_257		Stool	Malawi
This study	28099_1_258		Stool	Malawi
This study	28099_1_26		Stool	Malawi
This study	28099_1_260		Stool	Malawi
This study	28099_1_261		Stool	Malawi
This study	28099_1_264		Stool	Malawi
This study	28099_1_266		Stool	Malawi
This study	28099_1_268		Stool	Malawi
This study	28099_1_269		Stool	Malawi
This study	28099_1_27		Stool	Malawi
This study	28099_1_270		Stool	Malawi
This study	28099_1_272		Stool	Malawi
This study	28099_1_273		Stool	Malawi
This study	28099_1_274		Stool	Malawi
This study	28099_1_277		Stool	Malawi
This study	28099_1_278		Stool	Malawi

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
This study	28099_1_280		Stool	Malawi
This study	28099_1_281		Stool	Malawi
This study	28099_1_282		Stool	Malawi
This study	28099_1_284		Stool	Malawi
This study	28099_1_285		Stool	Malawi
This study	28099_1_286		Stool	Malawi
This study	28099_1_288		Stool	Malawi
This study	28099_1_289		Stool	Malawi
This study	28099_1_293		Stool	Malawi
This study	28099_1_294		Stool	Malawi
This study	28099_1_297		Stool	Malawi
This study	28099_1_30		Stool	Malawi
This study	28099_1_300		Stool	Malawi
This study	28099_1_301		Stool	Malawi
This study	28099_1_302		Stool	Malawi
This study	28099_1_303		Stool	Malawi
This study	28099_1_305		Stool	Malawi
This study	28099_1_306		Stool	Malawi
This study	28099_1_307		Stool	Malawi
This study	28099_1_309		Stool	Malawi
This study	28099_1_31		Stool	Malawi
This study	28099_1_311		Stool	Malawi
This study	28099_1_313		Stool	Malawi
This study	28099_1_314		Stool	Malawi
This study	28099_1_315		Stool	Malawi
This study	28099_1_317		Stool	Malawi
This study	28099_1_318		Stool	Malawi
This study	28099_1_319		Stool	Malawi
This study	28099_1_321		Stool	Malawi
This study	28099_1_322		Stool	Malawi
This study	28099_1_323		Stool	Malawi
This study	28099_1_325		Stool	Malawi

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
This study	28099_1_326		Stool	Malawi
This study	28099_1_327		Stool	Malawi
This study	28099_1_329		Stool	Malawi
This study	28099_1_330		Stool	Malawi
This study	28099_1_331		Stool	Malawi
This study	28099_1_333		Stool	Malawi
This study	28099_1_334		Stool	Malawi
This study	28099_1_335		Stool	Malawi
This study	28099_1_337		Stool	Malawi
This study	28099_1_338		Stool	Malawi
This study	28099_1_339		Stool	Malawi
This study	28099_1_34		Stool	Malawi
This study	28099_1_341		Stool	Malawi
This study	28099_1_342		Stool	Malawi
This study	28099_1_343		Stool	Malawi
This study	28099_1_345		Stool	Malawi
This study	28099_1_346		Stool	Malawi
This study	28099_1_347		Stool	Malawi
This study	28099_1_349		Stool	Malawi
This study	28099_1_35		Stool	Malawi
This study	28099_1_350		Stool	Malawi
This study	28099_1_351		Stool	Malawi
This study	28099_1_353		Stool	Malawi
This study	28099_1_354		Stool	Malawi
This study	28099_1_355		Stool	Malawi
This study	28099_1_357		Stool	Malawi
This study	28099_1_358		Stool	Malawi
This study	28099_1_359		Stool	Malawi
This study	28099_1_361		Stool	Malawi
This study	28099_1_362		Stool	Malawi
This study	28099_1_363		Stool	Malawi

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
This study	28099_1_365		Stool	Malawi
This study	28099_1_366		Stool	Malawi
This study	28099_1_367		Stool	Malawi
This study	28099_1_370		Stool	Malawi
This study	28099_1_371		Stool	Malawi
This study	28099_1_373		Stool	Malawi
This study	28099_1_374		Stool	Malawi
This study	28099_1_375		Stool	Malawi
This study	28099_1_377		Stool	Malawi
This study	28099_1_378		Stool	Malawi
This study	28099_1_379		Stool	Malawi
This study	28099_1_38		Stool	Malawi
This study	28099_1_381		Stool	Malawi
This study	28099_1_382		Stool	Malawi
This study	28099_1_383		Stool	Malawi
This study	28099_1_39		Stool	Malawi
This study	28099_1_41		Stool	Malawi
This study	28099_1_42		Stool	Malawi
This study	28099_1_43		Stool	Malawi
This study	28099_1_46		Stool	Malawi
This study	28099_1_47		Stool	Malawi
This study	28099_1_49		Stool	Malawi
This study	28099_1_50		Stool	Malawi
This study	28099_1_51		Stool	Malawi
This study	28099_1_53		Stool	Malawi
This study	28099_1_54		Stool	Malawi
This study	28099_1_55		Stool	Malawi
This study	28099_1_57		Stool	Malawi
This study	28099_1_58		Stool	Malawi
This study	28099_1_59		Stool	Malawi
This study	28099_1_61		Stool	Malawi
This study	28099_1_62		Stool	Malawi

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
This study	28099_1_63		Stool	Malawi
This study	28099_1_65		Stool	Malawi
This study	28099_1_66		Stool	Malawi
This study	28099_1_69		Stool	Malawi
This study	28099_1_7		Stool	Malawi
This study	28099_1_70		Stool	Malawi
This study	28099_1_71		Stool	Malawi
This study	28099_1_73		Stool	Malawi
This study	28099_1_74		Stool	Malawi
This study	28099_1_75		Stool	Malawi
This study	28099_1_77		Stool	Malawi
This study	28099_1_78		Stool	Malawi
This study	28099_1_79		Stool	Malawi
This study	28099_1_81		Stool	Malawi
This study	28099_1_82		Stool	Malawi
This study	28099_1_83		Stool	Malawi
This study	28099_1_85		Stool	Malawi
This study	28099_1_86		Stool	Malawi
This study	28099_1_87		Stool	Malawi
This study	28099_1_89		Stool	Malawi
This study	28099_1_90		Stool	Malawi
This study	28099_1_91		Stool	Malawi
This study	28099_1_93		Stool	Malawi
This study	28099_1_94		Stool	Malawi
This study	28099_1_95		Stool	Malawi
This study	28099_1_98		Stool	Malawi
This study	28099_1_99		Stool	Malawi
This study	28099_2_101		Stool	Malawi
This study	28099_2_105		Stool	Malawi
This study	28099_2_109		Stool	Malawi
This study	28099_2_113		Stool	Malawi
This study	28099_2_117		Stool	Malawi

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
This study	28099_2_12		Stool	Malawi
This study	28099_2_121		Stool	Malawi
This study	28099_2_126		Stool	Malawi
This study	28099_2_130		Stool	Malawi
This study	28099_2_138		Stool	Malawi
This study	28099_2_142		Stool	Malawi
This study	28099_2_146		Stool	Malawi
This study	28099_2_150		Stool	Malawi
This study	28099_2_154		Stool	Malawi
This study	28099_2_158		Stool	Malawi
This study	28099_2_16		Stool	Malawi
This study	28099_2_162		Stool	Malawi
This study	28099_2_166		Stool	Malawi
This study	28099_2_170		Stool	Malawi
This study	28099_2_174		Stool	Malawi
This study	28099_2_178		Stool	Malawi
This study	28099_2_182		Stool	Malawi
This study	28099_2_186		Stool	Malawi
This study	28099_2_190		Stool	Malawi
This study	28099_2_194		Stool	Malawi
This study	28099_2_198		Stool	Malawi
This study	28099_2_206		Stool	Malawi
This study	28099_2_210		Stool	Malawi
This study	28099_2_215		Stool	Malawi
This study	28099_2_219		Stool	Malawi
This study	28099_2_223		Stool	Malawi
This study	28099_2_227		Stool	Malawi
This study	28099_2_231		Stool	Malawi
This study	28099_2_235		Stool	Malawi
This study	28099_2_239		Stool	Malawi
This study	28099_2_24		Stool	Malawi

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
This study	28099_2_243		Stool	Malawi
This study	28099_2_247		Stool	Malawi
This study	28099_2_251		Stool	Malawi
This study	28099_2_255		Stool	Malawi
This study	28099_2_259		Stool	Malawi
This study	28099_2_263		Stool	Malawi
This study	28099_2_283		Stool	Malawi
This study	28099_2_287		Stool	Malawi
This study	28099_2_291		Stool	Malawi
This study	28099_2_295		Stool	Malawi
This study	28099_2_299		Stool	Malawi
This study	28099_2_3		Stool	Malawi
This study	28099_2_304		Stool	Malawi
This study	28099_2_308		Stool	Malawi
This study	28099_2_316		Stool	Malawi
This study	28099_2_32		Stool	Malawi
This study	28099_2_320		Stool	Malawi
This study	28099_2_36		Stool	Malawi
This study	28099_2_40		Stool	Malawi
This study	28099_2_44		Stool	Malawi
This study	28099_2_48		Stool	Malawi
This study	28099_2_56		Stool	Malawi
This study	28099_2_60		Stool	Malawi
This study	28099_2_64		Stool	Malawi
This study	28099_2_76		Stool	Malawi
This study	28099_2_8		Stool	Malawi
This study	28099_2_80		Stool	Malawi
This study	28099_2_84		Stool	Malawi
This study	28099_2_88		Stool	Malawi
This study	28099_2_92		Stool	Malawi
This study	28099_2_96		Stool	Malawi
This study	28099_2_97		Stool	Malawi

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Ingle 2018	100269_aEPEC	ERR134513	Stool	Gambia
Ingle 2018	100383_aEPEC	ERR137807	Stool	Gambia
Ingle 2018	100446	ERR178176	Stool	Gambia
Ingle 2018	100554_aEPEC	ERR134514	Stool	Gambia
Ingle 2018	100600_aEPEC	ERR134515	Stool	Gambia
Ingle 2018	102010_aEPEC	ERR137808	Stool	Gambia
Ingle 2018	102014_aEPEC	ERR137809	Stool	Gambia
Ingle 2018	102298_aEPEC	ERR134516	Stool	Gambia
Ingle 2018	102328_aEPEC	ERR134517	Stool	Gambia
Ingle 2018	102366_aEPEC	ERR137810	Stool	Gambia
Ingle 2018	102485_aEPEC	ERR134518	Stool	Gambia
Ingle 2018	103151	ERR178192	Stool	Gambia
Ingle 2018	200135_aEPEC	ERR134519	Stool	Mali
Ingle 2018	200232	ERR178150	Stool	Mali
Ingle 2018	200439_aEPEC	ERR134520	Stool	Mali
Ingle 2018	200456_aEPEC	ERR137812	Stool	Mali
Ingle 2018	200499	ERR178148	Stool	Mali
Ingle 2018	200696	ERR178151	Stool	Mali
Ingle 2018	200708_aEPEC	ERR137782	Stool	Mali
Ingle 2018	200758_aEPEC	ERR137783	Stool	Mali
Ingle 2018	200781_aEPEC	ERR124658	Stool	Mali
Ingle 2018	200959_aEPEC	ERR137813	Stool	Mali
Ingle 2018	201191_aEPEC	ERR137814	Stool	Mali
Ingle 2018	201214_aEPEC	ERR134521	Stool	Mali
Ingle 2018	201350	ERR178216	Stool	Mali
Ingle 2018	201381_aEPEC	ERR137784	Stool	Mali
Ingle 2018	201488_aEPEC	ERR137815	Stool	Mali
Ingle 2018	201534_aEPEC	ERR134522	Stool	Mali
Ingle 2018	201589_aEPEC	ERR137816	Stool	Mali
Ingle 2018	202317_aEPEC	ERR137817	Stool	Mali
Ingle 2018	202374	ERR178152	Stool	Mali
Ingle 2018	202387	ERR178149	Stool	Mali

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Ingle 2018	202423_aEPEC	ERR134523	Stool	Mali
Ingle 2018	202443_aEPEC	ERR134524	Stool	Mali
Ingle 2018	202453_aEPEC	ERR134525	Stool	Mali
Ingle 2018	202474	ERR178153	Stool	Mali
Ingle 2018	202521_aEPEC	ERR124659	Stool	Mali
Ingle 2018	202621_aEPEC	ERR137818	Stool	Mali
Ingle 2018	202833_aEPEC	ERR134526	Stool	Mali
Ingle 2018	202973_aEPEC	ERR134527	Stool	Mali
Ingle 2018	203470_aEPEC	ERR124660	Stool	Mali
Ingle 2018	204263_aEPEC	ERR124661	Stool	Mali
Ingle 2018	204302_aEPEC	ERR134528	Stool	Mali
Ingle 2018	300073	ERR178193	Stool	Mozambique
Ingle 2018	300086_aEPEC	ERR134529	Stool	Mozambique
Ingle 2018	300711_aEPEC	ERR134530	Stool	Mozambique
Ingle 2018	300795_aEPEC	ERR134531	Stool	Mozambique
Ingle 2018	300812_aEPEC	ERR137819	Stool	Mozambique
Ingle 2018	300814_aEPEC	ERR137820	Stool	Mozambique
Ingle 2018	302082	ERR178198	Stool	Mozambique
Ingle 2018	302302	ERR178154	Stool	Mozambique
Ingle 2018	302613	ERR178210	Stool	Mozambique
Ingle 2018	302619	ERR178211	Stool	Mozambique
Ingle 2018	302700	ERR178217	Stool	Mozambique
Ingle 2018	302701	ERR178212	Stool	Mozambique
Ingle 2018	302710	ERR178218	Stool	Mozambique
Ingle 2018	400549_aEPEC	ERR137785	Stool	Kenya
Ingle 2018	400654_aEPEC	ERR137786	Stool	Kenya
Ingle 2018	400714_aEPEC	ERR137787	Stool	Kenya
Ingle 2018	400896	ERR178177	Stool	Kenya
Ingle 2018	400897_aEPEC	ERR137821	Stool	Kenya
Ingle 2018	400998_aEPEC	ERR137789	Stool	Kenya
Ingle 2018	401082	ERR178178	Stool	Kenya

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Ingle 2018	401117_aEPEC	ERR137822	Stool	Kenya
Ingle 2018	401174_aEPEC	ERR137823	Stool	Kenya
Ingle 2018	401250_aEPEC	ERR137790	Stool	Kenya
Ingle 2018	401352	ERR178199	Stool	Kenya
Ingle 2018	401363	ERR178179	Stool	Kenya
Ingle 2018	401480_aEPEC	ERR124657	Stool	Kenya
Ingle 2018	401553	ERR178155	Stool	Kenya
Ingle 2018	401596_aEPEC	ERR137791	Stool	Kenya
Ingle 2018	401686	ERR178200	Stool	Kenya
Ingle 2018	401709_aEPEC	ERR137824	Stool	Kenya
Ingle 2018	401886_aEPEC	ERR137792	Stool	Kenya
Ingle 2018	401907	ERR178201	Stool	Kenya
Ingle 2018	401938_aEPEC	ERR137793	Stool	Kenya
Ingle 2018	402048_aEPEC	ERR134532	Stool	Kenya
Ingle 2018	402058_aEPEC	ERR137825	Stool	Kenya
Ingle 2018	402074_aEPEC	ERR137794	Stool	Kenya
Ingle 2018	402078	ERR178180	Stool	Kenya
Ingle 2018	402097_aEPEC	ERR137826	Stool	Kenya
Ingle 2018	402099_aEPEC	ERR134533	Stool	Kenya
Ingle 2018	402138_aEPEC	ERR124662	Stool	Kenya
Ingle 2018	402227_aEPEC	ERR137827	Stool	Kenya
Ingle 2018	402248_aEPEC	ERR134534	Stool	Kenya
Ingle 2018	402403	ERR178181	Stool	Kenya
Ingle 2018	402480_aEPEC	ERR137795	Stool	Kenya
Ingle 2018	402605	ERR178194	Stool	Kenya
Ingle 2018	402617	ERR178156	Stool	Kenya
Ingle 2018	402635	ERR178157	Stool	Kenya
Ingle 2018	402654_aEPEC	ERR134535	Stool	Kenya
Ingle 2018	402696	ERR178202	Stool	Kenya
Ingle 2018	402743_aEPEC	ERR137796	Stool	Kenya
Ingle 2018	402767	ERR178203	Stool	Kenya
Ingle 2018	402770_aEPEC	ERR134536	Stool	Kenya

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Ingle 2018	402780_aEPEC	ERR137797	Stool	Kenya
Ingle 2018	402794	ERR178204	Stool	Kenya
Ingle 2018	402798	ERR178182	Stool	Kenya
Ingle 2018	402837	ERR178183	Stool	Kenya
Ingle 2018	402842	ERR178205	Stool	Kenya
Ingle 2018	402898	ERR178184	Stool	Kenya
Ingle 2018	402924	ERR178158	Stool	Kenya
Ingle 2018	402977_aEPEC	ERR137798	Stool	Kenya
Ingle 2018	403066	ERR178159	Stool	Kenya
Ingle 2018	403096_aEPEC	ERR137799	Stool	Kenya
Ingle 2018	403128_aEPEC	ERR134537	Stool	Kenya
Ingle 2018	403308_aEPEC	ERR134538	Stool	Kenya
Ingle 2018	403523	ERR178206	Stool	Kenya
Ingle 2018	403726_aEPEC	ERR137800	Stool	Kenya
Ingle 2018	403728	ERR178161	Stool	Kenya
Ingle 2018	500094	ERR178207	Stool	India
Ingle 2018	500095	ERR178208	Stool	India
Ingle 2018	500193	ERR178213	Stool	India
Ingle 2018	500197_aEPEC	ERR137828	Stool	India
Ingle 2018	500275_aEPEC	ERR134539	Stool	India
Ingle 2018	500618_aEPEC	ERR134540	Stool	India
Ingle 2018	500858_aEPEC	ERR134541	Stool	India
Ingle 2018	500864_aEPEC	ERR134542	Stool	India
Ingle 2018	500989	ERR178164	Stool	India
Ingle 2018	501016	ERR178195	Stool	India
Ingle 2018	501029_aEPEC	ERR134543	Stool	India
Ingle 2018	503023	ERR178196	Stool	India
Ingle 2018	503028_aEPEC	ERR134544	Stool	India
Ingle 2018	503130_aEPEC	ERR137829	Stool	India
Ingle 2018	503225_aEPEC	ERR134545	Stool	India
Ingle 2018	503238_aEPEC	ERR137801	Stool	India
Ingle 2018	503256	ERR178197	Stool	India

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Ingle 2018	503311_aEPEC	ERR124653	Stool	India
Ingle 2018	503320	ERR178219	Stool	India
Ingle 2018	503331	ERR178165	Stool	India
Ingle 2018	503459_aEPEC	ERR134546	Stool	India
Ingle 2018	503537_aEPEC	ERR124663	Stool	India
Ingle 2018	503662_aEPEC	ERR134547	Stool	India
Ingle 2018	503891_aEPEC	ERR137802	Stool	India
Ingle 2018	503947_aEPEC	ERR124654	Stool	India
Ingle 2018	504005_aEPEC	ERR137803	Stool	India
Ingle 2018	504180	ERR178166	Stool	India
Ingle 2018	504225_aEPEC	ERR134548	Stool	India
Ingle 2018	504300_aEPEC	ERR134549	Stool	India
Ingle 2018	504324	ERR178167	Stool	India
Ingle 2018	504449_aEPEC	ERR134550	Stool	India
Ingle 2018	504528	ERR178168	Stool	India
Ingle 2018	504647_aEPEC	ERR134551	Stool	India
Ingle 2018	504718	ERR178169	Stool	India
Ingle 2018	504821_aEPEC	ERR134552	Stool	India
Ingle 2018	504888_aEPEC	ERR134553	Stool	India
Ingle 2018	504925_aEPEC	ERR124664	Stool	India
Ingle 2018	505148	ERR178170	Stool	India
Ingle 2018	505393_aEPEC	ERR124655	Stool	India
Ingle 2018	505513_aEPEC	ERR124656	Stool	India
Ingle 2018	505545	ERR178171	Stool	India
Ingle 2018	602206	ERR178172	Stool	Bangladesh
Ingle 2018	602370_aEPEC	ERR134554	Stool	Bangladesh
Ingle 2018	700149	ERR178214	Stool	Pakistan
Ingle 2018	700337_aEPEC	ERR134555	Stool	Pakistan
Ingle 2018	700495_aEPEC	ERR134556	Stool	Pakistan
Ingle 2018	700851	ERR178173	Stool	Pakistan
Ingle 2018	700863	ERR178215	Stool	Pakistan

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Ingle 2018	702161_aEPEC	ERR134558	Stool	Pakistan
Ingle 2018	702328	ERR178174	Stool	Pakistan
Ingle 2018	702566	ERR178175	Stool	Pakistan
Ingle 2018	702745_aEPEC	ERR137804	Stool	Pakistan
Ingle 2018	702797	ERR178186	Stool	Pakistan
Ingle 2018	702890_aEPEC	ERR137805	Stool	Pakistan
Ingle 2018	702898_aEPEC	ERR137806	Stool	Pakistan
Ingle 2018	702971	ERR178185	Stool	Pakistan
Ingle 2018	703063	ERR178209	Stool	Pakistan
Ingle 2018	703108	ERR178187	Stool	Pakistan
Ingle 2018	703128	ERR178188	Stool	Pakistan
Ingle 2018	703258_aEPEC	ERR134559	Stool	Pakistan
Ingle 2018	703273	ERR178191	Stool	Pakistan
Ingle 2018	703753	ERR178189	Stool	Pakistan
Ingle 2018	703975_aEPEC	ERR134560	Stool	Pakistan
Ingle 2018	G100788-1A	ERR175731	Stool	Gambia
Ingle 2018	G302544	ERR178226	Stool	Mozambique
Ingle 2018	G302551	ERR178225	Stool	Mozambique
Ingle 2018	G303212	ERR175730	Stool	Mozambique
Ingle 2018	G400792	ERR175725	Stool	Kenya
Ingle 2018	G400871	ERR175724	Stool	Kenya
Ingle 2018	G401436	ERR175727	Stool	Kenya
Ingle 2018	G401529	ERR178227	Stool	Kenya
Ingle 2018	G500007	ERR178223	Stool	India
Ingle 2018	G500297-1	ERR175733	Stool	India
Ingle 2018	G500407	ERR178221	Stool	India
Ingle 2018	G500830	ERR175728	Stool	India
Ingle 2018	G503854	ERR178224	Stool	India
Ingle 2018	G504540	ERR178222	Stool	India
Ingle 2018	G603423	ERR178228	Stool	Bangladesh
Ingle 2018	G702074-1	ERR175734	Stool	Pakistan
Ingle 2018	G702074-2	ERR175735	Stool	Pakistan

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Ingle 2018	R203092-3A	ERR175736	Stool	Mali
Ingle 2018	R203092-3B	ERR175737	Stool	Mali
Ingle 2018	R302583-2A	ERR175738	Stool	Mozambique
Ingle 2018	R302583-2B	ERR175739	Stool	Mozambique
Ingle 2018	R402077	ERR175726	Stool	Kenya
Ingle 2018	R503696	ERR175729	Stool	India
Mentzer 2014	E_1003	ERR054711	Stool	Egypt
Mentzer 2014	E_1009	ERR054712	Stool	Egypt
Mentzer 2014	E_1018	ERR084463	Stool	Egypt
Mentzer 2014	E_1034	ERR052911	Stool	Egypt
Mentzer 2014	E_1057CFn	ERR119471	Stool	Egypt
Mentzer 2014	E_106	ERR054666	Stool	unknown
Mentzer 2014	E_1072CFn	ERR119472	Stool	Egypt
Mentzer 2014	E_1074	ERR052912	Stool	Egypt
Mentzer 2014	E_1085	ERR052913	Stool	Egypt
Mentzer 2014	E_1091	ERR052914	Stool	Egypt
Mentzer 2014	E_110	ERR054678	Stool	Bangladesh
Mentzer 2014	E_1101CFn	ERR119473	Stool	Egypt
Mentzer 2014	E_1102CFn	ERR119474	Stool	Egypt
Mentzer 2014	E_1111	ERR052915	Stool	Egypt
Mentzer 2014	E_1167CFn	ERR119475	Stool	Egypt
Mentzer 2014	E_1169CFn	ERR119476	Stool	Egypt
Mentzer 2014	E_1189	ERR052916	Stool	Egypt
Mentzer 2014	E_1193CFn	ERR119477	Stool	Egypt
Mentzer 2014	E_1242CFn	ERR119478	Stool	Egypt
Mentzer 2014	E_1245	ERR052917	Stool	Egypt
Mentzer 2014	E_1248CFn	ERR119479	Stool	Egypt
Mentzer 2014	E_1258CFn	ERR119480	Stool	Egypt
Mentzer 2014	E_126	ERR054679	Stool	unknown
Mentzer 2014	E_1264CFn	ERR119481	Stool	Egypt
Mentzer 2014	E_1281CFn	ERR119482	Stool	Egypt
Mentzer 2014	E_1282CFn	ERR119483	Stool	Egypt

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Mentzer 2014	E_1285CFn	ERR119484	Stool	Egypt
Mentzer 2014	E_1287	ERR052918	Stool	Egypt
Mentzer 2014	E_129	ERR054680	Stool	Zaire
Mentzer 2014	E_1298	ERR052919	Stool	Egypt
Mentzer 2014	E_1316	ERR161000	Stool	Nepal
Mentzer 2014	E_133	ERR054681	Stool	unknown
Mentzer 2014	E_1334	ERR084464	Stool	China
Mentzer 2014	E_135	ERR054682	Stool	unknown
Mentzer 2014	E_1352CFn	ERR119485	Stool	Egypt
Mentzer 2014	E_1355CFn	ERR119486	Stool	Egypt
Mentzer 2014	E_1356CFn	ERR119487	Stool	Egypt
Mentzer 2014	E_1360_sec	ERR178234	Stool	Tunisia
Mentzer 2014	E_1362CFn	ERR119489	Stool	Egypt
Mentzer 2014	E_1363	ERR084466	Stool	Egypt
Mentzer 2014	E_1365CFn	ERR119490	Stool	Egypt
Mentzer 2014	E_1373	ERR052920	Stool	Indonesia
Mentzer 2014	E_1392	ERR052921	Stool	Indonesia
Mentzer 2014	E_1398CFn	ERR119491	Stool	Indonesia
Mentzer 2014	E_1407CFn	ERR119492	Stool	Mexico
Mentzer 2014	E_141	ERR054683	Stool	Burma
Mentzer 2014	E_1429tiny	ERR217371	Stool	Venezuela
Mentzer 2014	E_143	ERR054684	Stool	Japan
Mentzer 2014	E_1432G	ERR164830	Stool	Venezuela
Mentzer 2014	E_1432w	ERR164829	Stool	Venezuela
Mentzer 2014	E_1433	ERR084468	Stool	Morocco
Mentzer 2014	E_1460	ERR084469	Stool	Indonesia
Mentzer 2014	E_151	ERR054685	Stool	Japan
Mentzer 2014	E_1524	ERR084470	Stool	Argentina
Mentzer 2014	E_1525CFn	ERR119496	Stool	Argentina
Mentzer 2014	E_1526CFn	ERR119497	Stool	Argentina
Mentzer 2014	E_1527	ERR084471	Stool	Argentina

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Mentzer 2014	E_1532CFn	ERR119498	Stool	Argentina
Mentzer 2014	E_1533CFn	ERR119499	Stool	Argentina
Mentzer 2014	E_1534CFn	ERR119500	Stool	Argentina
Mentzer 2014	E_1535	ERR084472	Stool	Argentina
Mentzer 2014	E_1541	ERR052922	Stool	Argentina
Mentzer 2014	E_1542CFn	ERR119501	Stool	Argentina
Mentzer 2014	E_1543CFn	ERR119502	Stool	Argentina
Mentzer 2014	E_1544CFn	ERR119503	Stool	Argentina
Mentzer 2014	E_1548	ERR052923	Stool	Argentina
Mentzer 2014	E_1556CFn	ERR119504	Stool	Argentina
Mentzer 2014	E_1561CFn	ERR119505	Stool	Argentina
Mentzer 2014	E_1564CFn	ERR119506	Stool	Argentina
Mentzer 2014	E_1573CFn	ERR119508	Stool	Argentina
Mentzer 2014	E_1574CFn	ERR119509	Stool	Argentina
Mentzer 2014	E_1576CFn	ERR119510	Stool	Argentina
Mentzer 2014	E_157CFn	ERR119507	Stool	Japan
Mentzer 2014	E_1580CFn	ERR119511	Stool	Argentina
Mentzer 2014	E_1581CFn	ERR119512	Stool	Argentina
Mentzer 2014	E_1582CFn	ERR119513	Stool	Argentina
Mentzer 2014	E_1585CFn	ERR119514	Stool	Argentina
Mentzer 2014	E_1586CFn	ERR119515	Stool	Argentina
Mentzer 2014	E_1587CFn	ERR119516	Stool	Argentina
Mentzer 2014	E_1592CFn	ERR119518	Stool	Argentina
Mentzer 2014	E_1593	ERR084473	Stool	Argentina
Mentzer 2014	E_1594CFn	ERR119519	Stool	Argentina
Mentzer 2014	E_1596CFn	ERR119520	Stool	Argentina
Mentzer 2014	E_1597CFn	ERR119521	Stool	Argentina
Mentzer 2014	E_1599CFn	ERR119522	Stool	Argentina
Mentzer 2014	E_159CFn	ERR119517	Stool	Japan
Mentzer 2014	E_1600CFn	ERR119524	Stool	Argentina
Mentzer 2014	E_1604CFn	ERR119525	Stool	Argentina
Mentzer 2014	E_1607CFn	ERR119526	Stool	Argentina

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Mentzer 2014	E_1609CFn	ERR119527	Stool	Argentina
Mentzer 2014	E_1611CFn	ERR119528	Stool	Argentina
Mentzer 2014	E_1615CFn	ERR119529	Stool	Argentina
Mentzer 2014	E_1616CFn	ERR119530	Stool	Argentina
Mentzer 2014	E_1617CFn	ERR119531	Stool	Argentina
Mentzer 2014	E_1620	ERR084474	Stool	Argentina
Mentzer 2014	E_1623CFn	ERR119532	Stool	Indonesia
Mentzer 2014	E_1624	ERR052924	Stool	Indonesia
Mentzer 2014	E_1625CFn	ERR119533	Stool	Indonesia
Mentzer 2014	E_1628CFn	ERR119534	Stool	Indonesia
Mentzer 2014	E_1634CFn	ERR119535	Stool	Indonesia
Mentzer 2014	E_1635	ERR084475	Stool	Indonesia
Mentzer 2014	E_1637CFn	ERR119536	Stool	Indonesia
Mentzer 2014	E_1638CFn	ERR119537	Stool	Indonesia
Mentzer 2014	E_1640CFn	ERR119538	Stool	Indonesia
Mentzer 2014	E_1641CFn	ERR119539	Stool	Indonesia
Mentzer 2014	E_1642CFn	ERR119540	Stool	Indonesia
Mentzer 2014	E_1646	ERR052925	Stool	Indonesia
Mentzer 2014	E_1647CFn	ERR119541	Stool	Indonesia
Mentzer 2014	E_1648CFn	ERR119542	Stool	Indonesia
Mentzer 2014	E_1649	ERR084476	Stool	Indonesia
Mentzer 2014	E_1650CFn	ERR119543	Stool	Indonesia
Mentzer 2014	E_1654	ERR054665	Stool	Indonesia
Mentzer 2014	E_1657	ERR084477	Stool	Indonesia
Mentzer 2014	E_1659CFn	ERR119544	Stool	Indonesia
Mentzer 2014	E_1661CFn	ERR119545	Stool	Indonesia
Mentzer 2014	E_1666CFn	ERR119546	Stool	Indonesia
Mentzer 2014	E_1667	ERR084478	Stool	Indonesia
Mentzer 2014	E_1673CFn	ERR119548	Stool	Indonesia
Mentzer 2014	E_1674CFn	ERR119549	Stool	Indonesia
Mentzer 2014	E_1679sec	ERR217373	Stool	Indonesia
Mentzer 2014	E_167CFn	ERR119547	Stool	Japan

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Mentzer 2014	E_1682CFn	ERR119551	Stool	Indonesia
Mentzer 2014	E_1684CFn	ERR119552	Stool	Indonesia
Mentzer 2014	E_1690CFn	ERR119553	Stool	Indonesia
Mentzer 2014	E_1712CFn	ERR119554	Stool	Bangladesh
Mentzer 2014	E_1716	ERR052926	Stool	Bangladesh
Mentzer 2014	E_1724	ERR084479	Stool	Bangladesh
Mentzer 2014	E_1735CFn	ERR119555	Stool	Bangladesh
Mentzer 2014	E_1736CFn	ERR119556	Stool	Bangladesh
Mentzer 2014	E_1739CFn	ERR119557	Stool	Bangladesh
Mentzer 2014	E_1741	ERR084480	Stool	Bangladesh
Mentzer 2014	E_1744CFn	ERR119558	Stool	Bangladesh
Mentzer 2014	E_1750	ERR084481	Stool	Bangladesh
Mentzer 2014	E_1752CFn	ERR119559	Stool	Bangladesh
Mentzer 2014	E_1760	ERR084482	Stool	Bangladesh
Mentzer 2014	E_1779	ERR052927	Stool	Bangladesh
Mentzer 2014	E_1784	ERR052928	Stool	Bangladesh
Mentzer 2014	E_1795	ERR084483	Stool	Bangladesh
Mentzer 2014	E_1797	ERR084484	Stool	Bangladesh
Mentzer 2014	E_1841	ERR084485	Stool	Bangladesh
Mentzer 2014	E_1871CFn	ERR119560	Stool	Bangladesh
Mentzer 2014	E_1883	ERR084486	Stool	Bangladesh
Mentzer 2014	E_1918	ERR084487	Stool	Bangladesh
Mentzer 2014	E_1939	ERR084488	Stool	Bangladesh
Mentzer 2014	E_1947	ERR084489	Stool	Bangladesh
Mentzer 2014	E_1961CFn	ERR119561	Stool	Bangladesh
Mentzer 2014	E_1994	ERR084490	Stool	Bangladesh
Mentzer 2014	E_2088	ERR084491	Stool	Bangladesh
Mentzer 2014	E_2092	ERR089723	Stool	Bangladesh
Mentzer 2014	E_21	ERR054667	Stool	Bangladesh
Mentzer 2014	E_2108CFn	ERR119562	Stool	Bangladesh
Mentzer 2014	E_2110CFn	ERR119563	Stool	Bangladesh

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Mentzer 2014	E_2118	ERR089724	Stool	Bangladesh
Mentzer 2014	E_2121CFn	ERR119564	Stool	Bangladesh
Mentzer 2014	E_2131	ERR089725	Stool	Bangladesh
Mentzer 2014	E_2185	ERR089726	Stool	Bolivia
Mentzer 2014	E_220	ERR054686	Stool	Japan
Mentzer 2014	E_2219	ERR089727	Stool	Bolivia
Mentzer 2014	E_222CFn	ERR119380	Stool	Japan
Mentzer 2014	E_223CFn	ERR119381	Stool	Japan
Mentzer 2014	E_224	ERR054687	Stool	Japan
Mentzer 2014	E_2256	ERR089728	Stool	Thailand
Mentzer 2014	E_2339	ERR089729	Stool	Bolivia
Mentzer 2014	E_2347	ERR089730	Stool	Bolivia
Mentzer 2014	E_2348	ERR089731	Stool	Bolivia
Mentzer 2014	E_2362_sec	ERR178236	Stool	Bolivia
Mentzer 2014	E_237	ERR054688	Stool	Japan
Mentzer 2014	E_2370sec	ERR217375	Stool	Japan
Mentzer 2014	E_2386	ERR089732	Stool	Bolivia
Mentzer 2014	E_2397	ERR089733	Stool	Bolivia
Mentzer 2014	E_239CFn	ERR119382	Stool	Japan
Mentzer 2014	E_2439	ERR164832	Stool	Bolivia
Mentzer 2014	E_251	ERR054689	Stool	Japan
Mentzer 2014	E_263CFn	ERR119383	Stool	Japan
Mentzer 2014	E_272	ERR054690	Stool	Japan
Mentzer 2014	E_28	ERR054668	Stool	Bangladesh
Mentzer 2014	E_2980	ERR089734	Stool	Bangladesh
Mentzer 2014	E_2981_sec	ERR178239	Stool	Bangladesh
Mentzer 2014	E_329CFn	ERR119384	Stool	Mexico
Mentzer 2014	E_330CFn	ERR119385	Stool	Mexico
Mentzer 2014	E_333	ERR049162	Stool	Mexico
Mentzer 2014	E_335CFn	ERR119386	Stool	Mexico
Mentzer 2014	E_336CFn	ERR119387	Stool	Mexico
Mentzer 2014	E_340CFn	ERR119388	Stool	Mexico

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Mentzer 2014	E_343CFn	ERR119389	Stool	Mexico
Mentzer 2014	E_344	ERR049163	Stool	Mexico
Mentzer 2014	E_351CFn	ERR119390	Stool	Mexico
Mentzer 2014	E_354CFn	ERR119391	Stool	Mexico
Mentzer 2014	E_356CFn	ERR119392	Stool	Mexico
Mentzer 2014	E_36	ERR054669	Stool	Bangladesh
Mentzer 2014	E_360CFn	ERR119393	Stool	Mexico
Mentzer 2014	E_361CFn	ERR119394	Stool	Mexico
Mentzer 2014	E_370	ERR049164	Stool	Guatemala
Mentzer 2014	E_390CFn	ERR119395	Stool	Guatemala
Mentzer 2014	E_391CFn	ERR119396	Stool	Guatemala
Mentzer 2014	E_399CFn	ERR119397	Stool	Guatemala
Mentzer 2014	E_405CFn	ERR119398	Stool	Guatemala
Mentzer 2014	E_415CFn	ERR119399	Stool	Guatemala
Mentzer 2014	E_416	ERR049165	Stool	Guatemala
Mentzer 2014	E_425CFn	ERR119400	Stool	Guatemala
Mentzer 2014	E_445CFn	ERR119401	Stool	Guatemala
Mentzer 2014	E_45	ERR054670	Stool	Bangladesh
Mentzer 2014	E_451CFn	ERR119402	Stool	Guatemala
Mentzer 2014	E_471	ERR049166	Stool	Guatemala
Mentzer 2014	E_5089	ERR164833	Stool	Bangladesh
Mentzer 2014	E_509	ERR178229	Stool	Mexico
Mentzer 2014	E_513CFn	ERR119404	Stool	Mexico
Mentzer 2014	E_517	ERR049167	Stool	Mexico
Mentzer 2014	E_519CFn	ERR119405	Stool	Mexico
Mentzer 2014	E_520CFn	ERR119406	Stool	Mexico
Mentzer 2014	E_523CFn	ERR119407	Stool	Mexico
Mentzer 2014	E_527CFn	ERR119408	Stool	Mexico
Mentzer 2014	E_528CFn	ERR119409	Stool	Mexico
Mentzer 2014	E_529CFn	ERR119410	Stool	Mexico
Mentzer 2014	E_54	ERR049158	Stool	Bangladesh
Mentzer 2014	E_554	ERR049168	Stool	Mexico

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Mentzer 2014	E_562	ERR049169	Stool	Mexico
Mentzer 2014	E_563	ERR049170	Stool	Mexico
Mentzer 2014	E_604CFn	ERR119411	Stool	Guatemala
Mentzer 2014	E_616CFn	ERR119412	Stool	Guatemala
Mentzer 2014	E_618CFn	ERR119413	Stool	Guatemala
Mentzer 2014	E_620CFn	ERR119414	Stool	Guatemala
Mentzer 2014	E_621	ERR178230	Stool	Guatemala
Mentzer 2014	E_622CFn	ERR119416	Stool	Guatemala
Mentzer 2014	E_626CFn	ERR119417	Stool	Guatemala
Mentzer 2014	E_628CFn	ERR119418	Stool	Guatemala
Mentzer 2014	E_632	ERR049171	Stool	Guatemala
Mentzer 2014	E_636	ERR049172	Stool	Guatemala
Mentzer 2014	E_645CFn	ERR119419	Stool	Guatemala
Mentzer 2014	E_655	ERR049173	Stool	Guatemala
Mentzer 2014	E_658CFn	ERR119420	Stool	Guatemala
Mentzer 2014	E_659CFn	ERR119421	Stool	Guatemala
Mentzer 2014	E_66	ERR054671	Stool	Bangladesh
Mentzer 2014	E_662CFn	ERR119422	Stool	Guatemala
Mentzer 2014	E_70	ERR049159	Stool	Bangladesh
Mentzer 2014	E_703CFn	ERR119423	Stool	Guatemala
Mentzer 2014	E_704CFn	ERR119424	Stool	Guatemala
Mentzer 2014	E_705CFn	ERR119425	Stool	Guatemala
Mentzer 2014	E_71	ERR049160	Stool	Bangladesh
Mentzer 2014	E_710	ERR178231	Stool	Guatemala
Mentzer 2014	E_74	ERR054672	Stool	Bangladesh
Mentzer 2014	E_79	ERR054673	Stool	Bangladesh
Mentzer 2014	E_8	ERR049156	Stool	Bangladesh
Mentzer 2014	E_806	ERR054691	Stool	Guatemala
Mentzer 2014	E_810	ERR054692	Stool	Guatemala
Mentzer 2014	E_811	ERR178232	Stool	Guatemala
Mentzer 2014	E_812	ERR054693	Stool	Guatemala

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Mentzer 2014	E_816	ERR054694	Stool	Guatemala
Mentzer 2014	E_818CFn	ERR119428	Stool	Guatemala
Mentzer 2014	E_819CFn	ERR119429	Stool	Guatemala
Mentzer 2014	E_821CFn	ERR119430	Stool	Guatemala
Mentzer 2014	E_822	ERR054695	Stool	Guatemala
Mentzer 2014	E_828CFn	ERR119431	Stool	Guatemala
Mentzer 2014	E_833CFn	ERR119432	Stool	Guatemala
Mentzer 2014	E_841CFn	ERR119433	Stool	Guatemala
Mentzer 2014	E_842	ERR054696	Stool	Guatemala
Mentzer 2014	E_85	ERR054674	Stool	Bangladesh
Mentzer 2014	E_855CFn	ERR119434	Stool	Guatemala
Mentzer 2014	E_856	ERR054697	Stool	Guatemala
Mentzer 2014	E_858	ERR054698	Stool	Guatemala
Mentzer 2014	E_860CFn	ERR119435	Stool	Guatemala
Mentzer 2014	E_863	ERR049174	Stool	Guatemala
Mentzer 2014	E_865CFn	ERR119436	Stool	Guatemala
Mentzer 2014	E_867	ERR054699	Stool	Guatemala
Mentzer 2014	E_87	ERR054675	Stool	Bangladesh
Mentzer 2014	E_871	ERR054700	Stool	Guatemala
Mentzer 2014	E_873CFn	ERR119437	Stool	Guatemala
Mentzer 2014	E_876CFn	ERR119438	Stool	Guatemala
Mentzer 2014	E_877	ERR049175	Stool	Guatemala
Mentzer 2014	E_879	ERR049176	Stool	Guatemala
Mentzer 2014	E_88	ERR049161	Stool	Bangladesh
Mentzer 2014	E_881CFn	ERR119439	Stool	Guatemala
Mentzer 2014	E_882	ERR054701	Stool	Guatemala
Mentzer 2014	E_883	ERR054702	Stool	Guatemala
Mentzer 2014	E_884CFn	ERR119441	Stool	Guatemala
Mentzer 2014	E_885CFn	ERR119442	Stool	Guatemala
Mentzer 2014	E_887CFn	ERR119443	Stool	Guatemala
Mentzer 2014	E_888CFn	ERR119444	Stool	Guatemala
Mentzer 2014	E_890CFn	ERR119445	Stool	Guatemala

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Mentzer 2014	E_891CFn	ERR119446	Stool	Guatemala
Mentzer 2014	E_892CFn	ERR119447	Stool	Guatemala
Mentzer 2014	E_893CFn	ERR119448	Stool	Guatemala
Mentzer 2014	E_895	ERR054703	Stool	Guatemala
Mentzer 2014	E_897	ERR049177	Stool	Guatemala
Mentzer 2014	E_898CFn	ERR119449	Stool	Guatemala
Mentzer 2014	E_899CFn	ERR119450	Stool	Guatemala
Mentzer 2014	E_900	ERR054704	Stool	Guatemala
Mentzer 2014	E_901	ERR054705	Stool	Guatemala
Mentzer 2014	E_903	ERR054706	Stool	Guatemala
Mentzer 2014	E_907	ERR054707	Stool	Guatemala
Mentzer 2014	E_908CFn	ERR119452	Stool	Guatemala
Mentzer 2014	E_916	ERR049178	Stool	Guatemala
Mentzer 2014	E_917	ERR049179	Stool	Guatemala
Mentzer 2014	E_920	ERR054708	Stool	Guatemala
Mentzer 2014	E_924CFn	ERR119453	Stool	Guatemala
Mentzer 2014	E_925	ERR052905	Stool	Guatemala
Mentzer 2014	E_927	ERR052906	Stool	Egypt
Mentzer 2014	E_928	ERR054709	Stool	Egypt
Mentzer 2014	E_934CFn	ERR119455	Stool	Egypt
Mentzer 2014	E_935CFn	ERR119456	Stool	Egypt
Mentzer 2014	E_936CFn	ERR119457	Stool	Egypt
Mentzer 2014	E_938	ERR052907	Stool	Egypt
Mentzer 2014	E_939CFn	ERR119458	Stool	Egypt
Mentzer 2014	E_940	ERR052908	Stool	Egypt
Mentzer 2014	E_941CFn	ERR119459	Stool	Egypt
Mentzer 2014	E_943	ERR054710	Stool	Egypt
Mentzer 2014	E_944CFn	ERR119460	Stool	Egypt
Mentzer 2014	E_945CFn	ERR119461	Stool	Egypt
Mentzer 2014	E_947CFn	ERR119462	Stool	Egypt
Mentzer 2014	E_949CFn	ERR119463	Stool	Egypt
Mentzer 2014	E_952CFn	ERR119464	Stool	Egypt

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Mentzer 2014	E_953	ERR052909	Stool	Egypt
Mentzer 2014	E_955CFn	ERR119465	Stool	Egypt
Mentzer 2014	E_956CFn	ERR119466	Stool	Egypt
Mentzer 2014	E_957	ERR164828	Stool	Egypt
Mentzer 2014	E_97	ERR054676	Stool	Bangladesh
Mentzer 2014	E_978CFn	ERR119468	Stool	Egypt
Mentzer 2014	E_986	ERR052910	Stool	Egypt
Mentzer 2014	E_99	ERR054677	Stool	Bangladesh
Mentzer 2014	E_995	ERR160999	Stool	Egypt
Mentzer 2014	E_996CFn	ERR119469	Stool	Egypt
Mentzer 2014	E_998CFn	ERR119470	Stool	Egypt
Mentzer 2014	E160CFn	ERR119523	Stool	Japan
Mentzer 2014	E2367CFn	ERR119566	Stool	Bolivia
Mentzer 2014	E2371CFn	ERR119568	Stool	Bolivia
Mentzer 2014	E2377CFn	ERR119569	Stool	Bolivia
Mentzer 2014	E2388CFn	ERR119570	Stool	Bolivia
Mentzer 2014	E2392CFn	ERR119571	Stool	Bolivia
Mentzer 2014	E2393CFn	ERR119572	Stool	Bolivia
Mentzer 2014	E2395CFn	ERR119573	Stool	Bolivia
Mentzer 2014	E2404CFn	ERR119574	Stool	Bolivia
Mentzer 2014	E2405CFn	ERR119575	Stool	Bolivia
Mentzer 2014	E3015CFn	ERR119577	Stool	Egypt
Mentzer 2014	E4134CFn	ERR119578	Stool	Israel
Mentzer 2014	E5049	ERR089738	Stool	India
Mentzer 2014	E5051	ERR089739	Stool	India
Mentzer 2014	E5052	ERR089740	Stool	India
Mentzer 2014	E5080	ERR089741	Stool	Bangladesh
Mentzer 2014	E5081	ERR089742	Stool	Bangladesh
Mentzer 2014	E5082	ERR089743	Stool	Bangladesh
Mentzer 2014	E5084	ERR089744	Stool	Bangladesh
Mentzer 2014	E5085	ERR089745	Stool	Bangladesh

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Mentzer 2014	E5086	ERR089746	Stool	Bangladesh
Mentzer 2014	E5087	ERR089747	Stool	Bangladesh
Mentzer 2014	E5088	ERR089748	Stool	Bangladesh
Mentzer 2014	ILBEcoli5442571	ERR279354	Stool	Kenya
Mentzer 2014	ILBEcoli5442572	ERR279355	Stool	Kenya
Mentzer 2014	ILBEcoli5442573	ERR279356	Stool	Kenya
Mentzer 2014	ILBEcoli5442574	ERR279357	Stool	Kenya
Mentzer 2014	ILBEcoli5442575	ERR279358	Stool	Kenya
Mentzer 2014	ILBEcoli5442576	ERR279359	Stool	Kenya
Mentzer 2014	ILBEcoli5442577	ERR279360	Stool	Kenya
Mentzer 2014	ILBEcoli5442578	ERR279361	Stool	Kenya
Mentzer 2014	ILBEcoli5442579	ERR279362	Stool	Kenya
Mentzer 2014	ILBEcoli5442580	ERR279363	Stool	Kenya
Mentzer 2014	ILBEcoli5442581	ERR279364	Stool	Kenya
Mentzer 2014	ILBEcoli5442582	ERR279365	Stool	Kenya
Mentzer 2014	ILBEcoli5442583	ERR279366	Stool	Kenya
Mentzer 2014	ILBEcoli5442587	ERR279370	Stool	Guinea Bissau
Mentzer 2014	ILBEcoli5442588	ERR279371	Stool	Guinea Bissau
Mentzer 2014	ILBEcoli5442589	ERR279372	Stool	Guinea Bissau
Mentzer 2014	ILBEcoli5442590	ERR279373	Stool	Guinea Bissau
Musicha 2017	3487STDY6036382	ERR926351	Blood	Malawi
Musicha 2017	3487STDY6036383	ERR926352	Blood	Malawi
Musicha 2017	3487STDY6036384	ERR926353	Blood	Malawi
Musicha 2017	3487STDY6036385	ERR926354	CSF	Malawi
Musicha 2017	3487STDY6036386	ERR926355	Blood	Malawi
Musicha 2017	3487STDY6036387	ERR926356	Blood	Malawi
Musicha 2017	3487STDY6036388	ERR926357	Blood	Malawi
Musicha 2017	3487STDY6036389	ERR926358	Blood	Malawi
Musicha 2017	3487STDY6036390	ERR926359	CSF	Malawi
Musicha 2017	3487STDY6036391	ERR926360	Blood	Malawi
Musicha 2017	3487STDY6036392	ERR926361	Blood	Malawi
Musicha 2017	3487STDY6036393	ERR926362	Blood	Malawi

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Musicha 2017	3487STDY6036394	ERR926363	Blood	Malawi
Musicha 2017	3487STDY6036395	ERR926364	Blood	Malawi
Musicha 2017	3487STDY6036396	ERR926365	CSF	Malawi
Musicha 2017	3487STDY6036397	ERR926366	Blood	Malawi
Musicha 2017	3487STDY6036398	ERR926367	Blood	Malawi
Musicha 2017	3487STDY6036399	ERR926368	Blood	Malawi
Musicha 2017	3487STDY6036400	ERR926369	CSF	Malawi
Musicha 2017	3487STDY6036403	ERR926372	CSF	Malawi
Musicha 2017	3487STDY6036404	ERR926373	Blood	Malawi
Musicha 2017	3487STDY6036405	ERR926374	CSF	Malawi
Musicha 2017	3487STDY6036406	ERR926375	Blood	Malawi
Musicha 2017	3487STDY6036407	ERR926376	CSF	Malawi
Musicha 2017	3487STDY6036408	ERR926377	Blood	Malawi
Musicha 2017	3487STDY6036409	ERR926378	Blood	Malawi
Musicha 2017	3487STDY6036410	ERR926379	Blood	Malawi
Musicha 2017	3487STDY6036411	ERR926380	Blood	Malawi
Musicha 2017	3487STDY6036412	ERR926381	Blood	Malawi
Musicha 2017	3487STDY6036413	ERR971988	CSF	Malawi
Musicha 2017	3487STDY6036414	ERR926382	Blood	Malawi
Musicha 2017	3487STDY6036415	ERR926383	CSF	Malawi
Musicha 2017	3487STDY6036416	ERR926384	CSF	Malawi
Musicha 2017	3487STDY6036417	ERR926385	CSF	Malawi
Musicha 2017	3487STDY6036418	ERR926386	CSF	Malawi
Musicha 2017	3487STDY6036420	ERR926388	CSF	Malawi
Musicha 2017	3487STDY6036421	ERR926389	Blood	Malawi
Musicha 2017	3487STDY6036422	ERR926390	CSF	Malawi
Musicha 2017	3487STDY6036423	ERR926391	CSF	Malawi
Musicha 2017	3487STDY6036424	ERR926392	Blood	Malawi
Musicha 2017	3487STDY6036425	ERR926393	Blood	Malawi
Musicha 2017	3487STDY6036426	ERR926394	Blood	Malawi
Musicha 2017	3487STDY6036427	ERR926395	Blood	Malawi
Musicha 2017	3487STDY6036428	ERR926396	Blood	Malawi

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Musicha 2017	3487STDY6036429	ERR926397	Blood	Malawi
Musicha 2017	3487STDY6036430	ERR926398	Blood	Malawi
Musicha 2017	3487STDY6036431	ERR926399	Blood	Malawi
Musicha 2017	3487STDY6036432	ERR926400	Blood	Malawi
Musicha 2017	3487STDY6036433	ERR926401	CSF	Malawi
Musicha 2017	3487STDY6036434	ERR926402	CSF	Malawi
Musicha 2017	3487STDY6036435	ERR926403	CSF	Malawi
Musicha 2017	3487STDY6036436	ERR926404	Blood	Malawi
Musicha 2017	3487STDY6036437	ERR926405	Blood	Malawi
Musicha 2017	3487STDY6036438	ERR926406	Blood	Malawi
Musicha 2017	3487STDY6036440	ERR926408	CSF	Malawi
Musicha 2017	3487STDY6036441	ERR926409	Blood	Malawi
Musicha 2017	3487STDY6036443	ERR926411	Blood	Malawi
Musicha 2017	3487STDY6036444	ERR926412	Blood	Malawi
Musicha 2017	3487STDY6036445	ERR926413	RS	Malawi
Musicha 2017	3487STDY6036446	ERR971989	Blood	Malawi
Musicha 2017	3487STDY6036447	ERR926414	CSF	Malawi
Musicha 2017	3487STDY6036448	ERR926415	Blood	Malawi
Musicha 2017	3487STDY6036449	ERR926416	Blood	Malawi
Musicha 2017	3487STDY6036450	ERR926417	RS	Malawi
Musicha 2017	3487STDY6036451	ERR926418	CSF	Malawi
Musicha 2017	3487STDY6036452	ERR926419	Blood	Malawi
Musicha 2017	3487STDY6036453	ERR926420	Blood	Malawi
Musicha 2017	3487STDY6036454	ERR971990	Blood	Malawi
Musicha 2017	3487STDY6036455	ERR926421	CSF	Malawi
Musicha 2017	3487STDY6036456	ERR971991	Blood	Malawi
Musicha 2017	3487STDY6036457	ERR926422	RS	Malawi
Musicha 2017	3487STDY6036458	ERR926423	Blood	Malawi
Musicha 2017	3487STDY6036460	ERR926425	CSF	Malawi
Musicha 2017	3487STDY6036461	ERR926426	Blood	Malawi
Musicha 2017	3487STDY6036462	ERR971992	Blood	Malawi

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Musicha 2017	3487STDY6036463	ERR926427	CSF	Malawi
Musicha 2017	3487STDY6036464	ERR926428	Blood	Malawi
Musicha 2017	3487STDY6036465	ERR926429	Blood	Malawi
Musicha 2017	3487STDY6036466	ERR926430	RS	Malawi
Musicha 2017	3487STDY6036467	ERR926431	RS	Malawi
Musicha 2017	3487STDY6036468	ERR926432	Blood	Malawi
Musicha 2017	3487STDY6036470	ERR926434	Blood	Malawi
Musicha 2017	3487STDY6036471	ERR926435	RS	Malawi
Musicha 2017	3487STDY6036473	ERR971994	RS	Malawi
Musicha 2017	3487STDY6036486	ERR926444	Blood	Malawi
Musicha 2017	3487STDY6036506	ERR971962	Blood	Malawi
Musicha 2017	3487STDY6036508	ERR971963	Blood	Malawi
Musicha 2017	3487STDY6036519	ERR972008	Blood	Malawi
Musicha 2017	3487STDY6036520	ERR971966	Blood	Malawi
Musicha 2017	3487STDY6036526	ERR971968	Blood	Malawi
Musicha 2017	3487STDY6036533	ERR971970	Blood	Malawi
Musicha 2017	3487STDY6036547	ERR971979	Blood	Malawi
Musicha 2017	3487STDY6036565	ERR971987	RS	Malawi
Runchaeron 2017	3898STDY6199571	ERR1218581	Urine	Thailand
Runchaeron 2017	3898STDY6199572	ERR1218582	Urine	Thailand
Runchaeron 2017	3898STDY6199573	ERR1218583	Urine	Thailand
Runchaeron 2017	3898STDY6199574	ERR1218584	Sputum	Thailand
Runchaeron 2017	3898STDY6199575	ERR1218585	Urine	Thailand
Runchaeron 2017	3898STDY6199576	ERR1218534	Blood	Thailand
Runchaeron 2017	3898STDY6199577	ERR1218586	Pus	Thailand
Runchaeron 2017	3898STDY6199578	ERR1218587	Blood	Thailand
Runchaeron 2017	3898STDY6199579	ERR1218588	Urine	Thailand
Runchaeron 2017	3898STDY6199580	ERR1218589	Urine	Thailand
Runchaeron 2017	3898STDY6199581	ERR1218590	Pus	Thailand
Runchaeron 2017	3898STDY6199582	ERR1218591	Pus	Thailand
Runchaeron 2017	3898STDY6199583	ERR1218592	Urine	Thailand
Runchaeron 2017	3898STDY6199584	ERR1218593	Urine	Thailand

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Runchaeron 2017	3898STDY6199585	ERR1218594	Urine	Thailand
Runchaeron 2017	3898STDY6199586	ERR1218595	Blood	Thailand
Runchaeron 2017	3898STDY6199587	ERR1218596	Pus	Thailand
Runchaeron 2017	3898STDY6199588	ERR1218597	Urine	Thailand
Runchaeron 2017	3898STDY6199589	ERR1218535	Blood	Thailand
Runchaeron 2017	3898STDY6199590	ERR1218598	Urine	Thailand
Runchaeron 2017	3898STDY6199591	ERR1218599	Pus	Thailand
Runchaeron 2017	3898STDY6199592	ERR1218600	Urine	Thailand
Runchaeron 2017	3898STDY6199593	ERR1218536	Blood	Thailand
Runchaeron 2017	3898STDY6199594	ERR1218601	Urine	Thailand
Runchaeron 2017	3898STDY6199595	ERR1218602	Blood	Thailand
Runchaeron 2017	3898STDY6199596	ERR1218537	Pus	Thailand
Runchaeron 2017	3898STDY6199597	ERR1218538	Pus	Thailand
Runchaeron 2017	3898STDY6199598	ERR1218603	Pus	Thailand
Runchaeron 2017	3898STDY6199599	ERR1218539	Urine	Thailand
Runchaeron 2017	3898STDY6199600	ERR1218604	Pus	Thailand
Runchaeron 2017	3898STDY6199601	ERR1218540	Pus	Thailand
Runchaeron 2017	3898STDY6199602	ERR1218605	Urine	Thailand
Runchaeron 2017	3898STDY6199603	ERR1218606	Urine	Thailand
Runchaeron 2017	3898STDY6199604	ERR1218607	Urine	Thailand
Runchaeron 2017	3898STDY6199605	ERR1218608	Blood	Thailand
Runchaeron 2017	3898STDY6199606	ERR1218609	Pus	Thailand
Runchaeron 2017	3898STDY6199607	ERR1218610	Pus	Thailand
Runchaeron 2017	3898STDY6199608	ERR1218541	Urine	Thailand
Runchaeron 2017	3898STDY6199609	ERR1218611	Blood	Thailand
Runchaeron 2017	3898STDY6199610	ERR1218542	Pus	Thailand
Runchaeron 2017	3898STDY6199611	ERR1218612	Urine	Thailand
Runchaeron 2017	3898STDY6199612	ERR1218613	Pus	Thailand
Runchaeron 2017	3898STDY6199613	ERR1218614	Blood	Thailand
Runchaeron 2017	3898STDY6199614	ERR1218543	Urine	Thailand
Runchaeron 2017	3898STDY6199615	ERR1218615	Blood	Thailand
Runchaeron 2017	3898STDY6199616	ERR1218616	Urine	Thailand

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Runchaeron 2017	3898STDY6199617	ERR1218617	Blood	Thailand
Runchaeron 2017	3898STDY6199618	ERR1218618	Blood	Thailand
Runchaeron 2017	3898STDY6199619	ERR1218619	Urine	Thailand
Runchaeron 2017	3898STDY6199620	ERR1218620	Urine	Thailand
Runchaeron 2017	3898STDY6199621	ERR1218621	Blood	Thailand
Runchaeron 2017	3898STDY6199622	ERR1218544	Urine	Thailand
Runchaeron 2017	3898STDY6199623	ERR1218622	Urine	Thailand
Runchaeron 2017	3898STDY6199624	ERR1218545	Blood	Thailand
Runchaeron 2017	3898STDY6199625	ERR1218623	Urine	Thailand
Runchaeron 2017	3898STDY6199626	ERR1218624	Urine	Thailand
Runchaeron 2017	3898STDY6199627	ERR1218625	Urine	Thailand
Runchaeron 2017	3898STDY6199628	ERR1218626	Urine	Thailand
Runchaeron 2017	3898STDY6199629	ERR1218627	Urine	Thailand
Runchaeron 2017	3898STDY6199630	ERR1218628	Blood	Thailand
Runchaeron 2017	3898STDY6199631	ERR1218629	Blood	Thailand
Runchaeron 2017	3898STDY6199632	ERR1218630	Pus	Thailand
Runchaeron 2017	3898STDY6199633	ERR1218631	Pus	Thailand
Runchaeron 2017	3898STDY6199634	ERR1218632	Urine	Thailand
Runchaeron 2017	3898STDY6199635	ERR1218633	Blood	Thailand
Runchaeron 2017	3898STDY6199636	ERR1218634	Pus	Thailand
Runchaeron 2017	3898STDY6199637	ERR1218635	Urine	Thailand
Runchaeron 2017	3898STDY6199638	ERR1218546	Urine	Thailand
Runchaeron 2017	3898STDY6199639	ERR1218636	Pus	Thailand
Runchaeron 2017	3898STDY6199640	ERR1218637	Urine	Thailand
Runchaeron 2017	3898STDY6199642	ERR1218639	Blood	Thailand
Runchaeron 2017	3898STDY6199643	ERR1218640	Pus	Thailand
Runchaeron 2017	3898STDY6199644	ERR1218641	Blood	Thailand
Runchaeron 2017	3898STDY6199645	ERR1218642	Canal	Thailand
Runchaeron 2017	3898STDY6199648	ERR1218643	Canal	Thailand
Runchaeron 2017	3898STDY6199649	ERR1218644	Canal	Thailand
Runchaeron 2017	3898STDY6199651	ERR1218549	Canal	Thailand

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Runchaeron 2017	3898STDY6199653	ERR1218551	Canal	Thailand
Runchaeron 2017	3898STDY6199654	ERR1218646	Canal	Thailand
Runchaeron 2017	3898STDY6199656	ERR1218552	Canal	Thailand
Runchaeron 2017	3898STDY6199657	ERR1218648	Canal	Thailand
Runchaeron 2017	3898STDY6199658	ERR1218553	Canal	Thailand
Runchaeron 2017	3898STDY6199659	ERR1218649	Canal	Thailand
Runchaeron 2017	3898STDY6199660	ERR1218650	Canal	Thailand
Runchaeron 2017	3898STDY6199661	ERR1218651	Canal	Thailand
Runchaeron 2017	3898STDY6199662	ERR1218652	Canal	Thailand
Runchaeron 2017	3898STDY6199664	ERR1218654	Canal	Thailand
Runchaeron 2017	3898STDY6199665	ERR1218655	Canal	Thailand
Runchaeron 2017	3898STDY6199667	ERR1218656	Canal	Thailand
Runchaeron 2017	3898STDY6199669	ERR1218658	Canal	Thailand
Runchaeron 2017	3898STDY6199670	ERR1218659	Canal	Thailand
Runchaeron 2017	3898STDY6199671	ERR1218660	Canal	Thailand
Runchaeron 2017	3898STDY6199672	ERR1218661	Canal	Thailand
Runchaeron 2017	3898STDY6199673	ERR1218662	Canal	Thailand
Runchaeron 2017	3898STDY6199674	ERR1218663	Canal	Thailand
Runchaeron 2017	3898STDY6199675	ERR1218664	Canal	Thailand
Runchaeron 2017	3898STDY6199677	ERR1218666	Canal	Thailand
Runchaeron 2017	3898STDY6199680	ERR1218669	Canal	Thailand
Runchaeron 2017	3898STDY6199682	ERR1218671	Canal	Thailand
Runchaeron 2017	3898STDY6199685	ERR1218674	Canal	Thailand
Runchaeron 2017	3898STDY6199686	ERR1218675	Canal	Thailand
Runchaeron 2017	3898STDY6199687	ERR1218676	Canal	Thailand
Runchaeron 2017	3898STDY6199689	ERR1218678	Canal	Thailand
Runchaeron 2017	3898STDY6199692	ERR1218681	Canal	Thailand
Runchaeron 2017	3898STDY6199693	ERR1218682	Canal	Thailand
Runchaeron 2017	3898STDY6199694	ERR1218683	Canal	Thailand
Runchaeron 2017	3898STDY6199695	ERR1218684	Canal	Thailand
Runchaeron 2017	3898STDY6199696	ERR1218685	Canal	Thailand
Runchaeron 2017	3898STDY6199697	ERR1218686	Canal	Thailand

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Runchaeron 2017	3898STDY6199698	ERR1218554	Canal	Thailand
Runchaeron 2017	3898STDY6199700	ERR1218688	Canal	Thailand
Runchaeron 2017	3898STDY6199701	ERR1218689	Canal	Thailand
Runchaeron 2017	3898STDY6199702	ERR1218690	Untreated hospital sewage	Thailand
Runchaeron 2017	3898STDY6199704	ERR1218692	Untreated hospital sewage	Thailand
Runchaeron 2017	3898STDY6199705	ERR1218693	Untreated hospital sewage	Thailand
Runchaeron 2017	3898STDY6199706	ERR1218694	Canal	Thailand
Runchaeron 2017	3898STDY6199707	ERR1218695	Canal	Thailand
Runchaeron 2017	3898STDY6199708	ERR1218696	Canal	Thailand
Runchaeron 2017	3898STDY6199709	ERR1218697	Farm	Thailand
Runchaeron 2017	3898STDY6199710	ERR1218698	Farm	Thailand
Runchaeron 2017	3898STDY6199713	ERR1218701	Farm	Thailand
Runchaeron 2017	3898STDY6199714	ERR1218702	Farm	Thailand
Runchaeron 2017	3898STDY6199715	ERR1218703	Farm	Thailand
Runchaeron 2017	3898STDY6199764	ERR1218705	Urine	Thailand
Runchaeron 2017	3898STDY6199766	ERR1218556	Pus	Thailand
Runchaeron 2017	3898STDY6199768	ERR1218706	Urine	Thailand
Runchaeron 2017	3898STDY6199769	ERR1218707	Urine	Thailand
Runchaeron 2017	3898STDY6199772	ERR1218708	Pus	Thailand
Runchaeron 2017	3898STDY6199773	ERR1218557	Pus	Thailand
Runchaeron 2017	3898STDY6199778	ERR1218709	Canal	Thailand
Runchaeron 2017	3898STDY6199780	ERR1218710	Canal	Thailand
Runchaeron 2017	3898STDY6199781	ERR1218558	Canal	Thailand
Runchaeron 2017	3898STDY6199784	ERR1218711	Canal	Thailand
Runchaeron 2017	3898STDY6199790	ERR1218559	Urine	Thailand
Runchaeron 2017	3898STDY6199792	ERR1218712	Blood	Thailand
Runchaeron 2017	3898STDY6199793	ERR1218713	Urine	Thailand
Runchaeron 2017	3898STDY6199796	ERR1218560	Pus	Thailand
Runchaeron 2017	3898STDY6199798	ERR1218714	Blood	Thailand
Runchaeron 2017	3898STDY6199799	ERR1218715	Canal	Thailand
Runchaeron 2017	3898STDY6199802	ERR1218716	Canal	Thailand
Runchaeron 2017	3898STDY6199804	ERR1218561	Canal	Thailand

Table 7.1: Details of samples included in global phylogeny (*continued*)

Study	Sample ID	Accession No.	Source	Country
Runchaeron 2017	3898STDY6199805	ERR1218717	Canal	Thailand
Runchaeron 2017	3898STDY6199806	ERR1218718	Canal	Thailand
Runchaeron 2017	3898STDY6199807	ERR1218719	Canal	Thailand
Runchaeron 2017	3898STDY6199808	ERR1218562	Canal	Thailand
Runchaeron 2017	3898STDY6199809	ERR1218720	Canal	Thailand
Runchaeron 2017	3898STDY6199815	ERR1218564	Farm	Thailand
Runchaeron 2017	3898STDY6199816	ERR1218723	Canal	Thailand
Runchaeron 2017	3898STDY6199923	ERR1218774	Canal	Thailand

Note:

RS = Rectal swab

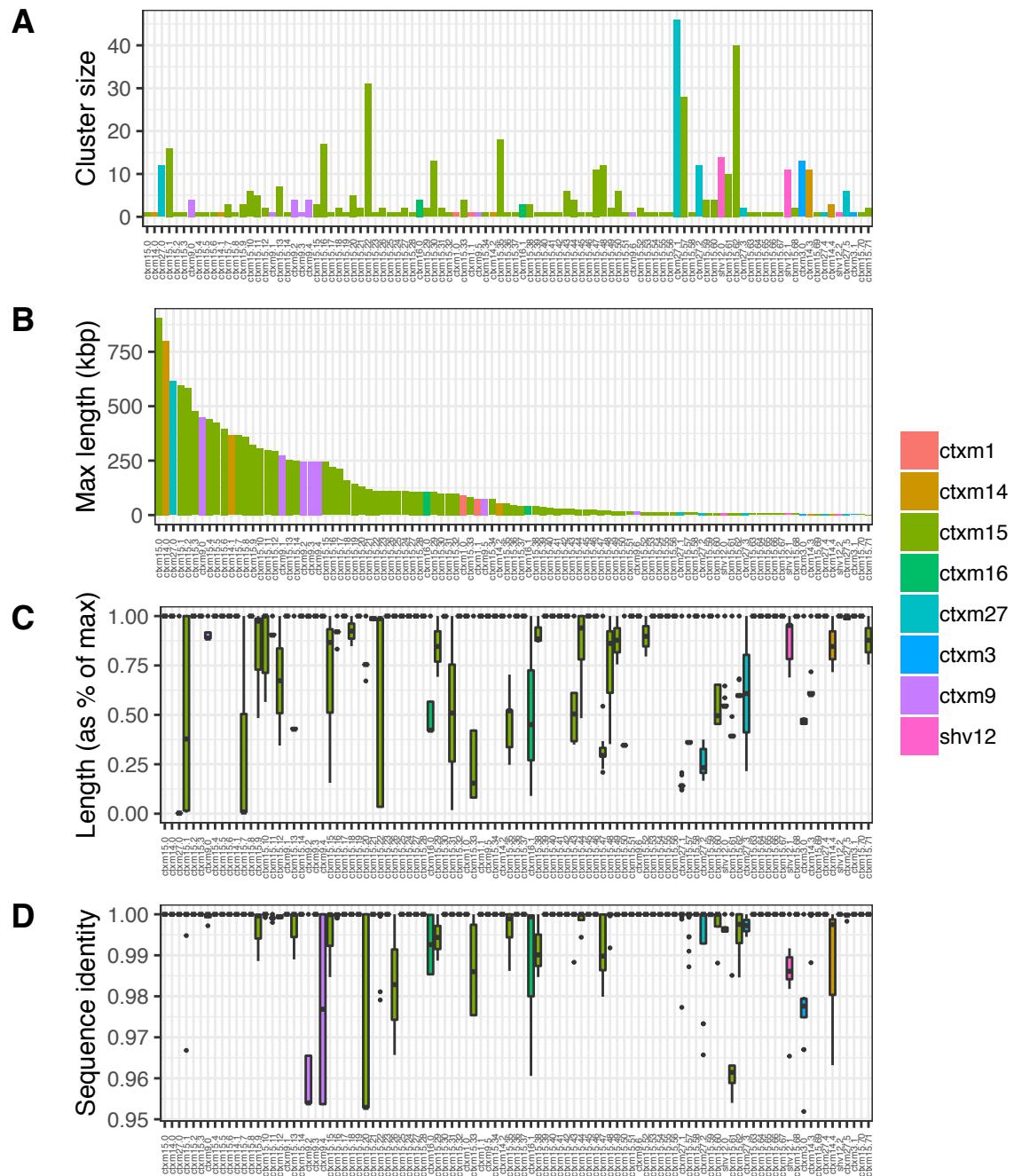


Figure 7.17: Summary statistics for 99 ESBL-containing contig clusters as determined by *cd-hit*. A: Number of contigs per cluster. B: Length (kbp) of longest sample in each cluster. This is defined as the cluster representative sample by *cd-hit* to which all other samples are compared for the purposes of length and sequence identity. C: Distribution of contig lengths by cluster expressed as a proportion of longest contig length. D: Distribution of sequence identity of cluster members compared to representative member, by cluster.



Figure 7.18: AMR genes, insertion sequences (IS) and plasmid replicons identified in the representative contig of each contig cluster, stratified by by ESBL gene and ordered by number of samples of cluster. IS26 is very frequently associated with a 108bp fragment of *catB4* chloramphenicol resistance gene, shown as a red fragment within the green IS26 element. A: *blaCTXM15*, B: *blaCTXM27* , C: *blaSHV12*. Plots show furthest IS/AMR gene or plasmid replicon up to +/- 10,000bp from the ESBL gene of interest.



Figure 7.19: AMR genes, insertion sequences (IS) and plasmid replicons identified in the representative contig of each contig cluster, stratified by ESBL gene and ordered by number of samples in cluster. IS26 is very frequently associated with a 108bp fragment of *catB4* chloramphenicol resistance gene, shown as a red fragment within the green IS26 element. A: *blaCTXM14*, B: *blaCTXM9*, C: *blaCTXM3*, D: *blaCTXM16*, E: *blaCTXM1*. Plots show furthest IS/AMR gene or plasmid replicon up to +/- 10,000bp from the ESBL gene of interest.

References

- 1 Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology* 2014;15:R46. doi:10.1186/gb-2014-15-3-r46
- 2 Bankevich A, Nurk S, Antipov D *et al.* SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology* 2012;19:455–77. doi:10.1089/cmb.2012.0021
- 3 Gurevich A, Saveliev V, Vyahhi N *et al.* QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013;29:1072–5. doi:10.1093/bioinformatics/btt086
- 4 Parks DH, Imelfort M, Skennerton CT *et al.* CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome research* 2015;25:1043–55. doi:10.1101/gr.186072.114
- 5 Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–9. doi:10.1093/bioinformatics/btu153
- 6 Page AJ, Cummins CA, Hunt M *et al.* Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 2015;31:3691–3. doi:10.1093/bioinformatics/btv421
- 7 Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* 2013;30:772–80. doi:10.1093/molbev/mst010
- 8 Page AJ, Taylor B, Delaney AJ *et al.* SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microbial Genomics* 2016;2. doi:10.1099/mgen.0.000056
- 9 Nguyen L-T, Schmidt HA, Haeseler A von *et al.* IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* 2015;32:268–74. doi:10.1093/molbev/msu300
- 10 Yu G, Smith DK, Zhu H *et al.* Ggtree : An r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and*

Evolution 2017;8:28–36. doi:10.1111/2041-210X.12628

- 11 Hunt M, Mather AE, Sánchez-Busó L *et al.* ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microbial genomics* 2017;3:e000131. doi:10.1099/mgen.0.000131
- 12 Inouye M, Dashnow H, Raven L-A *et al.* SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Medicine* 2014;6:90. doi:10.1186/s13073-014-0090-6
- 13 Carattoli A, Zankari E, García-Fernández A *et al.* In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrobial agents and chemotherapy* 2014;58:3895–903. doi:10.1128/AAC.02412-14
- 14 Clermont O, Christenson JK, Denamur E *et al.* The Clermont <i>Escherichia coli</i> phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. *Environmental Microbiology Reports* 2013;5:58–65. doi:10.1111/1758-2229.12019
- 15 Cheng L, Connor TR, Siren J *et al.* Hierarchical and Spatially Explicit Clustering of DNA Sequences with BAPS Software. *Molecular Biology and Evolution* 2013;30:1224–8. doi:10.1093/molbev/mst028
- 16 Altschul SF, Gish W, Miller W *et al.* Basic local alignment search tool. *Journal of Molecular Biology* 1990;215:403–10. doi:10.1016/S0022-2836(05)80360-2
- 17 Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 2006;22:1658–9. doi:10.1093/bioinformatics/btl158
- 18 Runcharoen C, Raven KE, Reuter S *et al.* Whole genome sequencing of ESBL-producing *Escherichia coli* isolated from patients, farm waste and canals in Thailand. *Genome Medicine* 2017;9:81. doi:10.1186/s13073-017-0471-8
- 19 Mentzer A von, Connor TR, Wieler LH *et al.* Identification of enterotoxigenic *Escherichia coli* (ETEC) clades with long-term global distribution. *Nature Genetics* 2014;46:1321–6. doi:10.1038/ng.3145
- 20 Ingle DJ, Tauschek M, Edwards DJ *et al.* Evolution of atypical enteropathogenic *E. coli* by repeated acquisition of LEE pathogenicity island variants. *Nature Microbiology* 2016;1:15010. doi:10.1038/nmicrobiol.2015.10
- 21 Musicha P, Feasey NA, Cain AK *et al.* Genomic landscape of extended-spectrum beta-lactamase resistance in *Escherichia coli* from an urban African setting. *J Antimicrob Chemother* 2017;72:1602–9. doi:10.1093/jac/dkx058

- 22 Dale AP, Woodford N. Extra-intestinal pathogenic Escherichia coli (ExPEC): Disease, carriage and clones. *Journal of Infection* 2015;71:615–26. doi:10.1016/J.JINF.2015.09.009
- 23 Krishnaraju M, Kamatchi C, Jha AK *et al.* Complete sequencing of an IncX3 plasmid carrying blaNDM-5 allele reveals an early stage in the dissemination of the blaNDM gene. *Indian journal of medical microbiology* 2015;33:30–8. doi:10.4103/0255-0857.148373
- 24 Ostrer L, Khodursky RF, Johnson JR *et al.* Analysis of mutational patterns in quinolone resistance-determining regions of GyrA and ParC of clinical isolates. *International Journal of Antimicrobial Agents* 2019;53:318–24. doi:10.1016/j.ijantimicag.2018.12.004
- 25 Ramirez MS, Tolmasky ME. Aminoglycoside modifying enzymes. *Drug resistance updates : reviews and commentaries in antimicrobial and anticancer chemotherapy* 2010;13:151–71. doi:10.1016/j.drup.2010.08.003
- 26 Galimand M, Courvalin P, Lambert T. Plasmid-Mediated High-Level Resistance to Aminoglycosides in Enterobacteriaceae Due to 16S rRNA Methylation. *Antimicrobial Agents and Chemotherapy* 2003;47:2565. doi:10.1128/AAC.47.8.2565-2571.2003
- 27 Anantham S, Hall RM. pCERC1, a Small, Globally Disseminated Plasmid Carrying the <i>dfrA14</i> Cassette in the <i>strA</i> Gene of the <i>sul2-strA-strB</i> Gene Cluster. *Microbial Drug Resistance* 2012;18:364–71. doi:10.1089/mdr.2012.0008
- 28 Nicolas-Chanoine M-H, Bertrand X, Madec J-Y. Escherichia coli ST131, an intriguing clonal group. *Clinical microbiology reviews* 2014;27:543–74. doi:10.1128/CMR.00125-13
- 29 Pitout JDD, DeVinney R. Escherichia coli ST131: a multidrug-resistant clone primed for global domination. *F1000Research* 2017;6:195. doi:10.12688/f1000research.10609.1
- 30 Stoesser N, Sheppard AE, Pankhurst L *et al.* Evolutionary History of the Global Emergence of the Escherichia coli Epidemic Clone ST131. *mBio* 2016;7:e02162. doi:10.1128/mBio.02162-15
- 31 Roer L, Overballe-Petersen S, Hansen F *et al.* <i>Escherichia coli</i> Sequence Type 410 Is Causing New International High-Risk Clones. *mSphere* 2018;3:e00337–18. doi:10.1128/mSphere.00337-18
- 32 Zong Z, Fenn S, Connor C *et al.* Complete genomic characterization of two Escherichia coli lineages responsible for a cluster of carbapenem-resistant infections in a Chinese hospital. *Journal of Antimicrobial Chemotherapy* 2018;73:2340–6. doi:10.1093/jac/dky210
- 33 Wang Q, Wang X, Wang J *et al.* Phenotypic and Genotypic Characterization of Carbapenem-resistant Enterobacteriaceae : Data From a Longitudinal Large-scale CRE Study in China (2012–2016). *Clinical Infectious Diseases* 2018;67:S196–205. doi:10.1093/cid/ciy660

- 34 Irrgang A, Falgenhauer L, Fischer J *et al.* CTX-M-15-Producing *E. coli* Isolates from Food Products in Germany Are Mainly Associated with an IncF-Type Plasmid and Belong to Two Predominant Clonal *E. coli* Lineages. *Frontiers in Microbiology* 2017;8:2318. doi:10.3389/fmicb.2017.02318
- 35 Kallonen T, Brodrick HJ, Harris SR *et al.* Systematic longitudinal survey of invasive <i>Escherichia coli</i> in England demonstrates a stable population structure only transiently disturbed by the emergence of ST131. *Genome Research* 2017;27:1437–49. doi:10.1101/gr.216606.116
- 36 Malawi Ministry of Health. Malawi Guidelines for Clinical Management of HIV in Children and Adults (Third Edition). 2016.
- 37 Neuert S, Nair S, Day MR *et al.* Prediction of Phenotypic Antimicrobial Resistance Profiles From Whole Genome Sequences of Non-typhoidal *Salmonella enterica*. *Frontiers in microbiology* 2018;9:592. doi:10.3389/fmicb.2018.00592
- 38 Williams CT, Musicha P, Feasey NA *et al.* ChloS-HRM, a novel assay to identify chloramphenicol-susceptible *Escherichia coli* and *Klebsiella pneumoniae* in Malawi. *The Journal of antimicrobial chemotherapy* 2019;74:1212. doi:10.1093/jac/dky563
- 39 Totsika M, Beatson SA, Sarkar S *et al.* Insights into a Multidrug Resistant *Escherichia coli* Pathogen of the Globally Disseminated ST131 Lineage: Genome Analysis and Virulence Mechanisms. *PLoS ONE* 2011;6:e26578. doi:10.1371/JOURNAL.PONE.0026578
- 40 Bevan ER, Jones AM, Hawkey PM. Global epidemiology of CTX-M β -lactamases: temporal and geographical shifts in genotype. *Journal of Antimicrobial Chemotherapy* 2017;72:2145–55. doi:10.1093/jac/dkx146
- 41 Musicha P, Cornick JE, Bar-Zeev N *et al.* Trends in antimicrobial resistance in bloodstream infection isolates at a large urban hospital in Malawi (1998–2016): a surveillance study. *The Lancet Infectious Diseases* 2017;17:1042–52. doi:10.1016/S1473-3099(17)30394-8
- 42 Duijkeren E van, Wielders CCH, Dierikx CM *et al.* Long-term Carriage of Extended-Spectrum β -Lactamase-Producing *Escherichia coli* and *Klebsiella pneumoniae* in the General Population in The Netherlands. *Clinical Infectious Diseases* 2018;66:1368–76. doi:10.1093/cid/cix1015
- 43 Stoesser N, Sheppard AE, Moore CE *et al.* Extensive Within-Host Diversity in Fe-cally Carried Extended-Spectrum-Beta-Lactamase-Producing *Escherichia coli* Isolates: Implications for Transmission Analyses. *Journal of clinical microbiology* 2015;53:2122–31. doi:10.1128/JCM.00378-15