Working title: "Comparative genomics and pangenomics of vancomycin resistant and susceptible *Enterococcus faecium* from Irish hospitals across 20 years"

Author names: Leigh R.J.[1], McKenna C.[1], McWade R.[2], Lynch B.[2], and Walsh F.[1]

Affiliations: [1]Department of Biology, Maynooth University

[2]Department of Microbiology, Mater Misericordiae University Hospital

Corresponding author: Dr. Robert Leigh (rob.leigh@mu.ie)

Keywords: Vancomycin resistance, Enterococcus, Evolution, Ireland, CC17

Repositories: Genome sequences and assembled genomes are deposited at (will be deposited to NCBI SRA and NCBI assembly upon publication).

Abstract:

Enterococcus faecium has emerged as an important nosocomial pathogen, which is increasingly difficult to treat due to the genetic acquisition of vancomycin resistance. Ireland exhibits a recalcitrant vancomycin resistant bloodstream infection rate compared to other developed countries. A set of 28 vancomycin resistant isolates was sequenced to construct a dataset alongside 61 other publicly available Irish genomes. This dataset was extensively analysed using *in-silico* methodologies and uncovered distinct evolutionary, coevolutionary, and clinically relevant population trends. These results suggest that a stable (in terms of genome size, GC%, and number of genes), yet genetically diverse population (in terms of gene

content) of *Enterococcus faecium* persist in Ireland with acquired resistance arising *via* plasmid acquisition (*vanA*) or to a lesser extent, chromosomal recombination (*vanB*). Population analysis described five clusters with one cluster partitioned into four clades which transcend isolation dates. Pangenomic and recombination analyses revealed an open (whole genome and chromosomal specific) pangenome illustrating a rampant evolutionary pattern. Comparative resistomics and virulomics uncovered distinct chromosomal and mobilomal propensity for multidrug resistance, widespread chromosomal point-mutation mediated resistance, and chromosomal harboured arsenals of virulence factors. Comparative phagomics revealed a core prophagome of three prophages throughout the dataset. Interestingly, a potential difference in biofilm formation strategies was highlighted by coevolutionary analysis, suggesting differential biofilm genotypes between *vanA* and *vanB* isolates. These results highlight the evolutionary history of Irish *Enterococcus faecium* isolates and may provide an insight into underlying infection dynamics in a clinical setting.

Introduction:

The genus *Enterococcus* (Firmicutes; Bacilli; Lactobacilli; Enterococcaceae) are commonly observed in diverse biomes such as soil, surface water, wastewater, and as commensal inhabitants of the higher chordate (inclusive of human) gastrointestinal (GI) tract, vaginal tract, and epidermis (1,2). Two species, *Enterococcus faecium* and *Enterococcus faecalis*, are aetiological of a cohort of moderate to severe conditions when migrated from the gastrointestinal tract, especially in the immunocompromised or convalescing host (3,4). Both *E. faecium* and *E. faecalis* migration can lead to caries, cellulitis, cholecystitis, cystitis, endocarditis, endodontits, periodontitis, peri-implantitis, postoperational peritonitis, sepsis, and neonatal meningitis (5–8). The clinical significance of these species was amplified as

antimicrobial resistance (AMR) began to evolve and disseminate throughout diverse environments (2,9).

Vancomycin is the drug-of-last-resort for recalcitrant infections with diverse resistance profiles (10), inclusive of the prominent nosocomial pathogen, methicillin-resistant *Staphylococcus aureus* (MRSA) and *E. faecium* (11,12). In recent years, vancomycin resistance has been observed in a number of human pathogens, often resulting in clinical complication and extended hospital stay (6,13–16). Ireland has one of the highest vancomycin resistant *E. faecium* (VRE) rates in Europe where 38.4% of blood stream infection isolates displaying resistance (ECDC: https://atlas.ecdc.europa.eu/). In particular, nosocomial VRE infection are of concern due to the rapid dissemination throughout this environment (4,15,17). Vancomycin resistance genes have been detected both on the chromosome and mobile elements (mobilome) and often accompanies other resistance genotypes (6,18–20). Commonly observed *vanA* mobilome co-resistance phenotypes are observed towards tetracycline, erythromycin, and aminoglycosides (21–24). Additionally, *E. faecium* almost ubiquitously displays chromosomal-mediated resistance to aminoglycosides (*via AAC*(6')-*Ii*), macrolides (*via efmA*, *efrAB* and *msrC*), and fluoroquinolones (*via gyrA* and *parC* mutations or upregulation of efflux), rifamycin (*via efrAB*), and to clindamycin, quinupristin-dalfopristin, and dalfopristin (*via lsaA*) (6,9,24–26).

In many countries, including the Republic of Ireland, vancomycin resistance is prominently derived from plasmid mediated *vanA*, and less commonly from chromosomal mediated *vanB* (2,27,28). Both *vanA* and *vanB* are D-Ala-D-Ala ligases (EC: 6.3.2.4) and confer resistance by constructing D-Ala-D-Lac as an alternative substrate during peptidoglycan synthesis, thus reducing vancomycin-peptidoglycan binding affinity (29–31). Aside from their associated genomic location (mobilomal *vs* chromosomal), *vanA* also confers resistance to

75 teicoplanin (another glycopeptide antibiotic) whereas *vanB* has not been observed to induce
76 this effect (32–34).

77 Metal and biocide resistance is of growing concern in a plethora of environments, and
78 especially in clinical settings (35). Correlations have been reported between metal resistance
79 and drug resistance in pathogenic bacteria, suggesting a potential co-evolutionary pressure on
80 both mechanisms (36–38). A range of metals are employed in healthcare for their intrinsic
81 antimicrobial properties, for example silver embedded plasters and copper plated door handles
82 (39–41). Previous studies have reported chromosomal mediated resistance to copper, silver,
83 selenium, and hydrogen peroxide in *E. faecium*, suggesting a growing resistance to passive
84 protection strategies (42–44).

85 Further to their resistance mechanisms, *Enterococcus* spp. employ a small cohort of
86 effective virulence factors during pathogenesis, allowing for adhesion, biofilm formation,
87 invasion, immunomodulation, and the synthesis of secreted toxins, enzymes, and peptides
88 (such as bacteriocins) into local environments to inhibit competition (9,45,46). Biofilm
89 formation is of critical importance in vancomycin resistant *Enterococcus* (VRE) infections,
90 due to difficulty of clearance and reduced antimicrobial penetration rates (42,47).

91 When isolated from human samples, the average *E. faecium* genome contains
92 2,765 ± 187 genes, yet the pangenome contained 12,457 when sampled from 161 genomes (26).
93 These results, and other pangenomic analyses (*e.g.* (48)) illustrate the genomic plasticity and
94 evolutionary capacity of *E. faecium*. While a major proportion of this variance is attributed to
95 mobile genetic elements (49), to our knowledge, no study has been conducted on chromosomal
96 pangenomic variance, so the extent of mobilome enticed variance is not fully elucidated. The
97 integration of phages (prophages) into bacterial chromosomes can introduce genetic novelty
98 (50–52). While a wide array of *Enterococcus* phages have been identified (53,54), and parasite-

host co-evolutionary trajectories have been explored (55), phage impact on symbiotic co-evolutionary trajectory has yet, to our knowledge, been explored.

The aims of this study were to compare the pan-genomes, mobilomes and chromosomes of the available genome sequences of vancomycin resistant or susceptible *E. faecium* across the timeframe that VRE (analysed by the ECDC: 2002 to 2019) increased in prevalence from 11.1% to 38.4% in Ireland (available at https://atlas.ecdc.europa.eu/public/index.aspx?Dataset=27&HealthTopic=4). Specifically, this study provides an insight into the pangenomics of *E. faecium* from three studies in Republic of Ireland using a comparative genomic, resistomic, and virulomic lens. It is now possible to derive intrapangenome correlations (Whelan *et al.*, 2020) so these approaches shall be explored to provide additional insight.

A recent study of the global dissemination of *E. faecium* identified that it has two main modes of genomic evolution: the acquisition and loss of genes, including antimicrobial resistance genes, through mobile genetic elements including plasmids, and homologous recombination of the core genome (49). Unfortunately, there were no Irish isolates contained within this study.

Methods:

Microbiological analysis

Twenty-eight isolates identified as vancomycin resistant *Enterococcus faecium* during 2018 and 2019 were collected by the *Mater Misericordiae* University Hospital (MMUH) in Dublin. The sample metadata are described in SI Table 1. Antimicrobial susceptibility testing was performed at the MMUH according to EUCAST guidelines

(https://www.eucast.org/ast_of_bacteria/) and subsequently at Maynooth University using disk testing according to the CLSI guidelines. The isolates were investigated for resistance to the following antibiotics: (ciprofloxacin, erythromycin, chloramphenicol, vancomycin, tetracycline, ampicillin, and linezolid) (SI Table 2).

DNA extractions genome sequencing

DNA was extracted from each of the 28 isolates using the Macherey-Nagel Nucleospin microbial DNA isolation kit according to the manufacturer's instructions. The extracted DNA was sequenced by Novogene using the "Bacterial resequencing" service on the Illumina with PE150 and Q30 ≥ 80%. This provided >100X coverage of each genome.

Genome assembly

The 28 samples sequenced for this study and all reads associated with Irish VRE genomes from British Society for Antimicrobial Chemotherapy study PRJEB4344 (hereafter referred to as the "BSAC" isolates (57)) were downloaded from the NCBI sequence read archive (SRA; (58)). Each read pair (sample) was subjected to adapter removal and quality trimming using TrimGalore *v.*0.6.6. (59) using default settings. Adapter removal during the TrimGalore pipeline was powered by CutAdapt *v.*3.0 (60) and FastQC *v.* 0.11.9 (61). Each sample was assembled using Unicycler *v.*0.4.7 (62) using default paired-end settings. Unicycler utilised SPAdes *v.*3.14.1. (63) to assemble reads and used Bowtie2 *v.*2.4.2. (64), Pilon *v.*1.23 (65), BLAST *v.*2.11 (66,67), and samtools *v.*1.11 (68) to further complete the assembly. Each assembly was quality assessed using CheckM (69) using the *Enterococcus faecium* database (SI Table 3) and sequence typed using MLST *v.*2.19.0 (Jolley and Maiden, 2010; Seemann,

2014) using default settings (SI Table 4; Figure 1). Instances where a ST could not be completely identified, the approximated alleles were used to approximate a ST using the "search by locus combinations" option for *E. faecium* using PubMLST (www.pubmlst.org (72)). Each assembly was separated into "chromosomes" and "plasmids" (hereafter referred to as "mobilomes" as complete plasmids were not always guaranteed and were treated collectively as an extrachromosomal entity) using Platon *v*.1.5.0 (73). These partitioned assemblies were analysed alongside their concatenated "whole genome" assemblies. Assemblies that had a reported completeness percentage ≤ 95% were retained for further analyses.

Additional genomes

A total of 11 *Enterococcus faecium* genome assemblies were available on NCBI assembly and attributed to Ireland in their respective metadata (from study PRJNA521309; (74)). As above, these assemblies were quality checked using CheckM, sequence typed using MLST, and separated into chromosomal and mobilomal components using Platon. These 11 isolates were all sampled in county Cork and shall be hereafter referred to as the "Cork" isolates.

Plasmid containment analysis

As plasmids were contiguous, a containment analysis was used to determine the closest relatives of the mobilome. Each containment analysis was performed using the "screen" algorithm in MASH *v.* 2.2.2. (75) against PLSDB (76) with a *P*-value stringency cut-off of 0.1

and an identity stringency cut-off of 0.99 to replicate the parameters used during the construction of PLSDB (Table SI 5).

Genome annotation

Each assembly dataset (whole genome, chromosomal, and mobilome) was annotated using Prokka *v.*1.14.6 (Seemann, 2014) using default settings. Prokka used Prodigal *v.*2.6.3. (78) to predict gene sequences, Aragorn *v.*1.2.38 (79) to detect tRNA sequences, and Minced *v.*0.4.0 (80) to detect CRISPR sequences, SignalP *v.*4.0 (81) to detect signal peptides, and HMMER *v.*3.3.1 for protein similarity searching (82), BioPerl *v.*1.7.2 for file manipulation (83), and barrnap *v.*0.9 for rRNA profiling (84). Each protein sequence in each assembly dataset was further annotated using InterProScan *v.*5.45-80.0 (85) using the "--appl PfamA" and "--goterms" to assign Pfam domains (86) and Gene Ontology terms (87) respectively. Finally, each assembly dataset was searched for secondary metabolite biosynthesis genes using a set of tools with their respective default settings (unless otherwise stated below): GECCO *v.*0.6.3 (88) and AntiSMASH *v.*5 (89), for transposable elements using MobileElementFinder *v.*1.0.3. (90), for antimicrobial resistance using ABRicate *v.*1.0.1 (Seemann, 2014) with the associated CARD database (92) and PointFinder *v.*1 (using the *Enterococcus faecium* database (93)), for virulence factors using Abricate with the associated VFDB dataset (94), and for metal (and biocide) resistance using BacMet *v.*2.0 (95). As BacMet is published as an amino acid dataset, it was first backtranslated to a representative nucleotide sequence (using translation table 11) using the "backtranseq" function in EMBOSS *v.*6.6.0.0 (96) with codon usage tables for *Enterococcus faecium* (the subject of this study); *Enterococcus faecalis*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumanii*, *Pseudomonas aeruginosa*, *Enterobacter* spp. (ESKAPE pathogens); *Escherichia coli* (model organism), *Treponema*

198 *pallidium*, *Neisseria gonorrhoeae*, *Chlamydia trachomatis* (prevalent bacterial STIs),
199 *Clostridium botulinum*, *Campylobacter jejuni*, *Listeria monocytogenes*, and *Vibrio*
200 *parahaemolyticus* (prominent foodbourne pathogens), *Helicobacter pylori* (common gastric
201 pathogen) and *Clostridiodes difficle*, *Legionella pneumophilia*, and *Mycobacterium*
202 *tuberculosis* (common nosocomial infections). Codon usage tables were obtained from the
203 Kazusa genome research institute (https://www.kazusa.or.jp/en/). To mitigate false negatives,
204 ABRicate was ran with a 50% minimum percentage identity stringency score (as opposed to
205 the default 80%) to allow for the detection of full-length homologs that may have been
206 otherwise undetected (Collated results for AntiSMASH and ABRicate are given in SI Tables
207 6-10).
208
209 Genome characteristic statistical analysis
210
211 The sum of coding genes, genome size (Mbp), genome density (the mean number of
212 genes per Mbp), and guanine-cytosine content (GC%) was calculated for the chromosomal and
213 mobilomal datasets for each sample (SI Table 11; Figure 2). Summary statistics (mean, median,
214 standard deviation, and variance) was computed for each data series (SI Table 12). Two-tailed
215 Welch's *t*-tests (H_0:$\mu_a=\mu_b$;H_A:$\mu_a\neq\mu_b$; (97,98)) were used to compare genome density and GC%
216 between chromosomal and mobilomal datasets, a Bonferroni-Dunn correction ($P_{BD} = P \times$
217 $n_{comparisons}$; $n_{comparisons} = 2$; (Bonferroni, 1936; Dunn, 1961)) was used to control Type-I errors
218 and instances where $P_{BD} \leq 0.005$ were considered statistically significant. A $P \leq 0.005$ will be
219 used to determine significance in all pairwise test comparisons to control for potential Type I
220 and Type II errors (101). Summary statistics were calculated for gene length in each genomic
221 subset (chromosomal and mobilomal), distributions were assessed for normality using a
222 Kolmogorov-Smirnoff test (H_0:~$X=N(\mu,\sigma)$;H_A:~$X\neq N(\mu,\sigma)$; (102,103)) and equivariance using

a Levene's test (H_0:$\sigma^2_a=\sigma^2_b$;H_A:$\sigma^2_a\neq\sigma^2_b$; (Levene, 1960)). All samples were observed to be non-normally distributed and 26 of 79 were observed to have unequal variances (Levene's test $P \leq 0.05$). Considering these trends, a Brunner-Munzel test (H_0:$B=0.5$;H_A:$B\neq0.5$) was used to determine whether chromosomes were statistically more likely to have longer genes than mobilomes. A Bonferroni-Dunn correction was applied ($n_{comparisons} = 79$) and instances where $P_{BD} \leq 0.005$ were considered statistically significant (SI Table 12).

Genome relatedness

Chromosomal sequences from each isolate were all-*vs*-all compared using MASH *v.* 2.2.2. (75) and instances where the reported distance (D) ≤ 0.05 with a P-value ≤ 0.05 extracted as an edge list and visualised as a network with Gephi *v.*0.9.2 (104) using the Fruchterman-Reingold algorithm (105) (Figure 3). This procedure was repeated for mobilomal sequences (Figure 4).

Pangenome analysis

A pangenome for the whole genome and chromosomal datasets (for all isolates, and individually for the "Cork" isolates, for the "BSAC" isolates, and for the isolates sequenced for this study) was produced using Roary *v.*3.1.3. (106) with the "-e" flag to align all gene clusters using PRANK *v.*170427 (107), and the "-z" flag to keep intermediate files (retained to produce the robust phylogeny below). To assess the effect of the mobilome on the genome diversity and complexity, each pangenomic category (n_{genes}; "core pangenome" ($99\% \leq n_{samples} \leq 100\%$), "soft core pangenome" ($95\% \leq n_{samples} < 99\%$), "shell pangenome" ($15\% \leq n_{samples} < 95\%$), and "cloud pangenome" ($n_{samples} < 15\%$)) in the whole genome generated pangenome

was compared to the chromosomal pangenome using a two-tailed Fisher's exact test (H_0:ρ=π;H_A:$\rho\neq\pi$; (108)) with a Bonferroni-Dunn correction ($n_{comparisons}$ = 4), instances where $P_{BD} \leq 0.05$ were considered statistically significant, and statistically significant instances where $\rho > \pi$ were considered to be overrepresented and instances where $\rho < \pi$ were considered underrepresented (SI Table 13).

Pangenome function

The representative pangenome refers to the collection of representative sequences from each pangenome cluster. Gene ontology terms (as assigned by InterProScan) were extracted and slimmed from the whole genome dataset (inclusive of all isolates) using the "map_to_slim.py" from GOATools (109) using generic .obo files. Sequences were grouped into their pangenomic categories and compared to the background population using the "find_enrichment.py" script (using a Fisher's exact test (H_0:ρ=π;H_A:$\rho\neq\pi$) and Bonferroni-Dunn correction (SI Table 14))

Phylogeny construction

A phylogeny was constructed using single-copy, ubiquitous gene alignments (as produced by PRANK during pangenome construction) from the chromosomal dataset. The chromosomal dataset was used to further minimise the likelihood of interference non-vertically inherited sequences on the phylogeny. Each alignment was quality trimmed using TrimAL *v.*1.4 (110) using the "-automated1" flag. A superalignment was constructed by concatenating all trimmed alignments using FASconCAT *v.*1.04 (111) and a consensus phylogeny was constructed using IQtree *v.*1.6.12 (112) with 10,000 bootstrap replicates. IQTree used

ModelTest-NG *v*.0.1.7 (113) to determine the GTR+F+R6 (114) nucleotide evolutionary model to be most appropriate for phylogenetic reconstruction. The root of the tree was determined by creating a second phylogeny with a specified outgroup of 2 *Staphylococcus aureus* genomes. The second phylogeny was constructed by clustering all chromosomal and staphylococcal proteins using ProteinOrtho *v*.6.0.24 (115) with a stringency cut off value of $E \leq 1.00e^{-50}$. All protein clusters that were ubiquitous in all *Enterococcus faecium* species and single copy for each species within each cluster were extracted and aligned using Muscle *v*.3.8.1551 (116). Each alignment was trimmed as above using TrimAL (using the "-automated1" flag) and a consensus tree was constructed as above using IQtree with 10,000 bootstrap replicates. ModelFinder-NG determined LG+I+G (117) to be the best model of protein evolution. The second phylogeny determined ST178 (all samples isolated in Cork, Ireland) to be the outgroup of the core gene tree. The finalised phylogeny was displayed and annotated using iToL *v*.5 (118) (Figure 5). The phylogeny was annotated with genome annotation data (generated using the analyses above and below) to visualise underlying trends (Figures 6-10; SI Figure 1)

Isolate clustering

Isolates were clustered using previously constructed core gene superalignment using RheirBAPS *v*.1.1.3 (119,120) with a maximum depth of two, an initial population cluster of 20, and with an additional 100 rounds of processing to approximate optimal clustering (Figure 5; Figure 11; SI Table 15). All isolate-cluster assignments were determined to approach 100% likelihood ($P = 1$ in all cases) indicating that their placement is correct.

Core genome MLST

298 A core genome MLST (cgMLST) was constructed for all isolates using chewBBACA
299 *v*.2.8.5 (121) using default settings and using the *Enterococcus faecium* Prodigal training file
300 provided with the software. The cgMLST profile was displayed and annotated with metadata
301 using GrapeTree (122) with the MSTree V2 algorithm (Figure 12). The generated cgMLST
302 was verified using our previously generated data pangenomic data with PANINI *v*.1 (123).
303 Briefly, as PANINI requires a minimum of 100 input taxa, we constructed a pseudo-dataset
304 where each taxon was represented twice and processed. Pseudo-taxa were removed from the
305 PANINI output and was visualised alongside the previously constructed phylogeny and
306 annotated with their associated RheirBAPS generated clade using MicroReact (124) (SI Figure
307 2).
308
309 Pangenomic co-evolutionary analyses
310
311 The chromosomal phylogeny and Roary-derived pangenome (inclusive of all isolates)
312 were used to test statistically significant associations or dissociations between any vancomycin
313 resistance gene (*vanA, vanB, vanH, vanW, vanX, vanXB, vanY, vanYB*) and any other gene (SI
314 Table 16) using Coinfinder (56). Coinfinder uses a binomial test ($H_0:\pi=\pi_x$;$H_A:\pi\neq\pi_x$) and while
315 the authors advise the use of a Bonferroni-Dunn corrected, this was not implemented so all
316 coincidences could be explored and instances where $P \leq 0.005$ were considered statistically
317 significant.
318
319 Recombination analysis
320
321 Using RheirBAPS resulted into two distinct clades, the "Cork" isolates (CC94) and all
322 other isolates (CC17; discussed in a later section). To examine the extent of core genome

recombination on CC17 evolution, genomic comparisons were performed between each CC17 isolate and a "complete genome" *Enterococcus faecium* (reference) assembly (downloaded from NCBI Assembly). The reference was selected from a pool of all available *Enterococcus faecium* complete genome assemblies (which passed taxonomy assignment) from NCBI Assembly, where each genome was assessed for *vanA* or *vanB* presence using ABRicate with the CARD database (--minid 50), assigned a sequence type using MLST, and assigned a CC using PubMLST. Any assembly with a *vanA*+ or *vanB*+ phenotype, with an indeterminant ST, or assigned to either CC17 or CC94 were discarded. Remaining isolates were processed to remove plasmids and the remaining chromosomal sequence was annotated using Prokka with default settings. Proteins from all CC17, CC94, and candidate reference assemblies were clustered using ProteinOrtho ($E \leq 1e^{-50}$). Protein clusters that were both single copy and ubiquitous in all species were individually aligned using Muscle with uninformative regions using TrimAL (with the "-automated1" flag) and concatenated into a superalignment using FASConCAT. The superalignment was processed to construct a consensus tree with 10,000 bootstrap replicates using IQTree. The resultant consensus tree was rooted at CC94 and the closest assembly to CC17 was selected as the reference strain. The most suitable reference was determined to be GCF_005166365.1 (strain NM213; PRJNA513159). Strain NM213 was isolated from healthy Egyptian infants in 2018 to determine its potential as a probiotic (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA513159/).

Whole genome alignments were constructed between strain NM213 and each CC17 assembly and concatenated using Snippy *v*.4.6.0 (snippy-multi) with default settings (https://github.com/tseemann/snippy). The resultant core whole multigenome alignment was processed to convert non-standard nucleotide characters to (ATGC) to "N" using a Snippy auxillary script ("snippy-clean_full_aln"). Genomic recombination regions were detected using Gubbins *v*.3.0.0 (125) using default settings and the extent of recombination was

illustrated using Phandango *v.*1.3.0 (126) where the previously constructed core-genome phylogeny (of CC17 and CC94) was used to scaffold recombination (Figure 13). This procedure was replicated for the remaining 24 suitable references (SI Figure 3).

Phages

Phage and prophage elements within each chromosomal and plasmid sequence were assessed using Phigaro *v.* 2.3.0 (127) with default settings and saving the detected viral sequences to their own output files ("--save-fasta" flag) (SI Table 17). The sum of phages per genome were counted (SI Table 18). Extracted viral sequences were assessed as a source of drug resistance, metal resistance and virulence factors using abricate with the CARD, BACMET, VFDB databases as above and as a source of secondary metabolism machinery using GECCO as above, only one gene-of-interest was observed (discussed below). Extracted prophages were annotated using Prokka using the "--kingdom Viruses" flag.

Phage classification

A database of all Caudoviridae ICTV exemplar reference phage genomes were downloaded from NCBI Assembly. Caudoviridae were selected as this clade encompassed all viral families identified by Phigaro (Siphoviridae, Inoviridae, and Myoviridae, respectively). Each extracted prophage was searched against the database using Mash *v.*2.2.2 (75) using the "dist" (distance) algorithm. As we searched prophages (partial genomes) against phages (whole genomes) we selected top hits (1 per prophage) from instances with the shortest distances ($D < 1$) and smallest *P*-values ($P \leq 0.005$) for each prophage (SI Table 18; Figure 14).

373 Integron assessment

374

375 The presence of integrons in chromosomal and plasmid sequences was determined
376 using IntegronFinder *v*.2.0 (128) using the "--local-max" flag and setting "--calin-threshold"
377 flag to 1 (instead of the default 2). We chose to set the "--calin-threshold" flag to 1 to include
378 integrated regions with CALIN artefacts. While lowering this threshold may yield a higher
379 number of false positives, we feel it is appropriate for exploratory purposes. The distribution
380 of integrons per genome is given in SI Table 19. Again, ABRicate was used to determine
381 whether contigs containing integron elements contained genes-of-interest (SI Table 20; Figure
382 15).

383

384 **Results**

385

386 Quality control

387

388 All assemblies were reported to have a high level of completeness ($\mu_{completeness}$ =
389 99.29±0.55%; 95.16 ≤ % ≤ 99.61; SI Table 3).

390

391 General characteristics of the genomes

392

393 Genomes used in this manuscript were reported to have a mean chromosomal gene
394 count of 2537.91±77.44 (η = 2550.5; 2334 ≤ n ≤ 2730) and an aggregated (chromosomal and
395 mobilomal) gene count of 2,821.64±116.65 (η = 2812; 2508 ≤ n ≤ 3158). For chromosomal
396 data, a mean gene sum of 2537.91±77.44, a mean genome size of 2681.5±78.77 Mbp, a mean
397 density of 0.946±0.0098 genes/Mbp, and a mean GC% of 38.03±0.15% (SI Tables 10-11;

Figure 2). For mobilomal data, a mean gene sum of 283.73±84.86, a mean genome size of 259.99±75.76 Mbp, a mean density of 1.088 genes/Mbp, and a mean GC% of 34.70±0.38%. When statistically examined, mobilomes were observed to be significantly denser with regards the number of observed genes (to have a greater number of genes per Mbp genome; $P_{BD} < 1.00e^{-06}$) and had a significantly lower GC% ($P_{BD} < 1.00e^{-06}$) When the coding proportions were compared, 19 mobilome samples were reported to be significantly less dense (in terms of the sum of coding nucleotides) than their chromosomal counterparts, however 61 comparisons displayed insignificant differences.

Genetic relatedness (MLST, cgMLST, and RheirBAPS)

While a total of 16 STs were observed throughout the dataset, considerable differences were observed between the 3 sources (isolates sequenced for this study, previously published isolates, and "Cork" isolates). The "Cork" isolates were exclusively ST178, and ST178 was not observed in either of the clinical sources (Figure 1, SI Table 4). Isolates sequenced for this study were predominantly ST80 (16 of 28; 57.14%) and previously published isolates were predominantly ST17 (17 of 40; 42.5%) and ST192 (9 of 40; 22.5%). Interestingly, ST80 and ST17 are exclusively observed in their associated sources, however ST192 and ST203 are observed in both sources.

Isolate clustering using RheirBAPS matched clonal complexes at level one, with all CC94 isolates assigned to Cluster A and all CC17 isolates assigned to Cluster B (SI Table 14). At level two, eight clusters were observed, however using phylogenetic inference, this can be collapsed to five clusters with four subclusters. As all CC94 isolates were again assigned to a single cluster, CC17 isolates formed four clusters (α, β, γ, δ) with four subclusters (δ_1, δ_2, δ_3, and δ_4). These clusters are named based on their phylogenetic divergences (Figure 5) . Cluster

α contained one genome, ERR374806 (ST1032) and was the earliest diverging CC17 isolate in our dataset. Cluster β encompassed all ST203 isolates and all ST192 isolates except for ERR374982 (which was assigned to Cluster γ, matching its phylogenetic placement). Cluster γ encompasses all ST78 isolates and the singleton ST192. As previously mentioned, Cluster δ was comprised of four distinct subclusters (δ_1, δ_2, δ_3, and δ_4). Where δ_1 is basal to all other δ clades, and δ_2 is basal to sister clades δ_3 and δ_4. Clade δ_1 encompassed all ST17 isolates, and isolates assigned to singleton ST groups (ST18, ST132, ST982, ST1038, ST1421). Clade δ_2 encompassed all ST80 isolates, two ST80 isolates, and singleton ST isolates (ST16, ST64, ST132, ST202, ST612). Clade δ_3 encompasses seven ST80 isolates and clade δ_4 encompasses the remaining eight ST80 isolates and the singleton ST787. These results suggest that ST80 is a descendant of ST17 and is undergoing considerable chromosomal evolution. Clusters α, β, and γ, and Subclusters δ_3, and δ_4 all displayed monophyly. Subcluster δ_2 displayed considerable paraphyly with 4 root nodes. Subcluster δ_1 also displayed paraphyly (if the Cluster δ earliest diverging taxon (Sample 8) is ignored) due to ERR374755. The Subcluster δ_1 paraphyly may be considered an artefact of cluster evolution, with Subclusters δ_2, δ_3, and δ_4 undergoing rapid evolution after the divergence of ERR374755 from their common ancestor. The paraphyly observed in Subcluster δ_2 may be due to sample size and may be resolved or further partitioned with additional taxa. The paraphyly observed is likely due to evolutionary expansion between sampling times, with earlier diverging isolates (BSAC isolates) being isolated prior to 2011 and later diverging isolates (isolates sequenced for this study) being sequenced in 2018).

The cgMLST graph (and associated graph analyses) displayed considerable modularity (Figure 12; SI Figure 2) with distinct clades forming, largely corroborating RheirBAPS results, regardless of the year of isolation. Two exceptions were observed for this trend: firstly, polyphyly in Cluster β was observed with isolates from 2009 adjacent to Cluster γ, and those from 2018 adjacent to Subcluster δ_2. Again, Subcluster δ_2 appeared less modular than other

Clusters (mirroring paraphyly observed in the phylogeny), but this was likely due to core genome evolution between sampling dates. A Subcluster δ_2 isolate was observed adjacent to Subcluster δ_4.

Plasmid containment

As expected, the majority of PLSDB hits were attributed to *E. faecium* (SI Table 5), however some other interesting observations were also noted. All eleven CC94 (Cork) isolates each displayed a hit to two *Listeria monocytogenes* strain CFSAN023459 plasmids (pCFSAN023459_01 and pCFSAN023459_02), eight CC94 isolates also returned a hit for *Enterococcus hirae* strain CQP3-9 plasmid pCQP3-9_1. Hits pertaining to *Bacillus cereus* plasmid pBC16 were observed across CC17, interestingly, all Cluster γ isolates except ERR375014 (ST78) and all Cluster β isolates except ERR374813 and ERR374914 (ST203; only Cluster β isolates not sequenced during this study) returned a hit. Plasmid pEGM182-2 from *Enterococcus casseliflavus* strain EGM182 was observed in CC94 and in Subclusters δ_3 and δ_4, where four of six δ_3 isolates (except Samples 6 and 12; ST80) returned two hits for pEGM182-2. Hits were returned for *Escherichia coli* strain UB-ESBL31 plasmid pESBL31 from isolates across the phylogeny. Of note, all Cluster γ isolates and five δ_4 isolates returned a hit for pESBL31. Most of these isolates (14 of 15) were sampled between 2002 and 2011, and one isolate (Sample 1) was sampled in 2018, suggesting persistent survivability for this plasmid and its descendants.

Antimicrobial susceptibility profiles

The 28 vancomycin resistant *Enterococcus faecium* were isolated from 22 hospital patients in 2018/2019 (MMUH, Dublin isolates). Samples 2 and 24 were isolated on the same day from patient 2. Samples 10, 5 and 27 were isolated from patient 5 at days 92, 100 and 107, respectively. Samples 21, 20, 9 and 26 were isolated from patient 9 on days 0, 8, 19 and 365, respectively. The antimicrobial susceptibility profiles (SI Table 2) of the isolates confirmed they were all vancomycin resistant. Of the 28 VRE*fm*, 25 were erythromycin resistant, two were chloramphenicol resistant, 28 were ampicillin resistant, 19 were tetracycline resistant and two were linezolid resistant. Samples 14 and 24 were the only isolates resistant to chloramphenicol or linezolid. While sample 2 was isolated from the same patient as sample 24 it did not display the same resistance profile. The three samples susceptible to erythromycin did not display a specific AMR gene absence associated only with these samples. The metadata of the BSAC study contained susceptibility to vancomycin data only and the "Cork" isolates contained no antimicrobial resistance data. Of the BSAC study isolates 23 were vancomycin susceptible and 17 were vancomycin resistant (SI Table 1).

Antimicrobial resistance genotyping

A total of 32 resistance genes were observed throughout the dataset, of which 13 were exclusively observed in the chromosome (AAC(6')-Ii, *dfrF*, *ermT*, *efmA*, *msrC*, *tetM*, *vanB*, *vanHB*, *vanRB*, *vanSB*, *vanWB*, *vanXB*, and *vanYB*), where one pair (*vanRB* and *vanSB*) were observed exclusively in isolate ERR374834. Ten genes were observed exclusively on the mobilome (ANT(6)-Ia, *Enterococcus faecium* chloramphenicol acetyltransferase, *catA8*, *lsaE*, *tetU*, *vanA*, *vanHA*, *vanRA*, *vanXA*, and *vanZA*), and 9 were observed in either the chromosome or mobilome (AAC(6')-Ie-APH(2")-Ia, APH(3')-IIIa, *ermB*, *SAT-4*, *aad(*6), *dfrG*, *tet(L)*, *vanSA*,

and *vanYA*). Interestingly, the 9 genes present on either the chromosome or mobilome never appeared both chromosomally and mobilomally in the same isolate (SI Table 6; Figure 6).

Mobile AMR genes

The *vanA* gene was present only in the mobilome of 44 isolates, almost always with the cluster of *vanHA*, *vanRA*, *vanSA*, *vanXA*, *vanYA* and *vanZA* (SI Table 6; Figure 6). These mobilomes were sampled from across CC17, spanning Clusters β, γ, and δ (encompassing 16 ST80 isolates, 7 ST78 isolates, 5 ST203 isolates, 4 ST17 isolates, and 3 ST192 genomes, and one isolate from each of ST16, ST18, ST64, ST132, ST202, ST612, ST1032, ST1421 and an unassigned ST).The *vanA* isolates comprised samples from the BSAC study (n = 16) isolated in 2002, 2003, 2005, 2006, 2007 and 2009 to 2011 and the isolates from Dublin between 2018 and 2019 (n = 28). The *vanB* gene was identified on the chromosomes in addition to the mobile *vanA* on two isolates from the BSAC study and one isolate from MMUH (sample 1). Samples 14 and 24 (ST80, Subcluster δ_4) were linezolid resistant but did not contain the mobile linezolid resistance genes. Both samples were also the only isolates displaying chloramphenicol resistance. However, neither contained any known mobile chloramphenicol resistance genes. *Enterococcus faecium* chloramphenicol acetyltransferase (conferring resistance to phenicols) was observed in the mobilome of 3 of 17 ST17 isolates and in the ST1038 isolate (all Subcluster δ_4). Another phenicol resistance gene *catA8* was also observed in the mobilome of sample 26. However, all were phenotypically susceptible to chloramphenicol.

The resistance gene *lsaE* (conferring resistance to lincosamide) was observed on the mobilomes of samples 18 and 26 and ten isolates from the BSAC study. The genes *aad(6)*, *ermB* and *aph(3')III-a* were present in each of the BSAC mobilomes containing *lsaE*. The tetracycline resistance gene *tetU* was observed on the mobilome of 6 ST80 isolates, 2 ST18

521 isolates, and one isolate from each of ST17, ST132, ST787, ST1038, and ST1421 respectively.
522 In addition, *tetL* (conferring resistance to tetracyclines) was frequently present on the
523 mobilomes.
524
525 Chromosomally mediated AMR
526
527 The *vanB* gene was identified in 11 isolates and were always chromosomal. A cluster
528 of *vanB, vanWB, vanXB,* and *vanYB* (conferring resistance to glycopeptides, specifically
529 vancomycin) was observed in nine Subcluster δ_1 isolates (accounting for one ST16, and 8 of
530 17 ST17 isolates) and one Subcluster δ_2 isolate (ERR374832; ST132) (SI Table 6; Figure 6).
531 With the exception of ST16, all isolates with these four genes were also observed to possess
532 *vanHB*, and the ST132 isolate was further observed to possess *vanRB* and *vanSB.* The genes
533 *vanSA* and *vanYA* were individually (and uniquely) observed in sample 8 (unassigned ST) and
534 sample 23 (ST203) respectively.
535 One chromosomal exclusive gene, *aac(6')-Ii* (aminoglycoside resistance), was
536 observed to be ubiquitous in all samples and one other *msrC* (macrolide and streptogramin B
537 resistance), was observed in all isolates except sample 23 (ST64). The combination of these
538 two genes is reported to confer resistance to compounds from the aminoglycoside, lincosamide,
539 macrolide, oxazolidinone, phenicol, pleuromutilin, streptogramin, and tetracycline classes
540 (129–132). Interestingly, CC94 isolates were not observed to possess any other AMR genes
541 beyond these two examples. The genes *ermB* and *ermT* (conferring resistance to lincosamide,
542 macrolide, and streptogramin) were present in 46 of 68 CC17 isolates. Only one isolate, sample
543 23 (ST64), was observed to possess both *ermB* and *ermT* genes and each individual gene was
544 distributed relatively unevenly throughout the dataset (*eg. ermB* was observed in 8 ST80
545 isolates, *ermT* in 5 ST80 isolates, and 5 ST80 isolates lacked both genes); however, it was

546 observed that, if present, an isolate only possessed one of *ermB* or *ermT* on the chromosome
547 (with the exception of Sample 23). The distribution of *erm* in this regard does seem to follow
548 an underlying phylogenetic bias. For example, approximately half of Subcluster δ_2 possess
549 *ermB* and the other possess *ermT*. This transition appears to have occurred after the divergence
550 of ERR374697 as it is retained in all later diverging taxa (Figures 5,6). Another example of
551 *erm* bias can be observed between Subclusters δ_2, δ_3, and δ_4, where δ_2, δ_3 display bias towards
552 *ermB* and δ_4 towards *ermT*. As both genes are observed in the Sample 23 chromosome (the
553 earliest diverging δ_4 taxon), it can be reasonably assumed that this transition occurred after the
554 divergence of Sample 23 from the last common ancestor of the remaining δ_4 isolates.

555 The gene *dfrF* (conferring resistance to trimethoprim) was observed distributed
556 throughout the dataset (like *ermBT*) and displayed phylogenetic biases towards Cluster γ, and
557 *ermT* positive members of Subclusters δ_1, δ_2, and δ_3. The gene *tetM* (conferring resistance to
558 tetracycline) was scattered throughout the dataset but was specifically observed in 13 or 18
559 ST80 isolates.

560

561 AMR genes present on either mobilome or chromosome.

562

563 Seven of the AMR genes detected were distributed across either the mobilome or the
564 chromosome within the samples investigated (*AAC(6')-Ie-APH(2'')-Ia*, *aph(3')-IIIa*, *ermB*,
565 *SAT*-4, *aad*(6), *dfrG* and *tetL*). This demonstrates the inter-connectedness of the mobilome and
566 the chromosome as a mode of AMR gene transport within these isolates (SI Table 6; Figure 6).

567

568 Core resistome

569

570 Using default pangenomic parameters (106), a core chromosomal resistome (inclusive
571 of soft-core) whereby a gene is represented in ≥95% (≥75 genomes) comprise of two genes:
572 *AAC(6')-Ii* and *msrC*. When CC94 is excluded, the core mobile resistome (≥64 genomes), is
573 expanded to include *efmA*. A core mobile resistome was not established.
574
575 Single nucleotide point mutations
576
577 Linezolid resistance was identified phenotypically in sister taxa, Samples 14 and 24,
578 but no associated plasmid mediated resistance mechanism was identified (SI Table 6; Figure
579 6). In addition, the 23S rRNA, L3 and L4 gene sequences and amino acid sequences,
580 respectively, were compared with the linezolid susceptible isolates in this study using
581 PointFinder (SI Table 7; Figure 7). No linezolid resistance associated mutations were
582 identified. Thus, the mechanism of linezolid resistance has not been identified. The linezolid
583 resistance phenotype of the BSAC or CC94 isolates was not reported.
584 The ciprofloxacin resistance phenotype was not reported in the BSAC or CC94 isolate
585 associated metadata. Isolates sequenced for this study were all ciprofloxacin resistant. The
586 gyrase and topoisomerase IV genes (*gyrA* and *parC*) from each isolate was compared with the
587 ciprofloxacin susceptible strains using PointFinder to identify point mutations associated with
588 ciprofloxacin resistance. Mutations in all isolates were identified at amino acid position 83 in
589 GyrA, resulting in either an S83I change in all δ_2 isolates (Samples 6, 7, 12, 15, 19, 20, and 22)
590 or an S83Y change in the remaining samples sequenced for this study, in *ermT* positive δ_1
591 isolates, both ST203 isolates, in the only Cluster α isolate, and in a four-isolate clade of δ_2. In
592 addition, two mutations at amino acid 709: Y709N and Y709D, outside the quinolone
593 resistance determining region (QRDR) in the same groups of isolates. Point mutations at p.S80
594 were observed across all CC17 isolates except ERR37474. The majority (53 of 78) of isolates

595 displayed an S80I mutation and the remaining 15 isolates displayed an S80R mutation. Point
596 mutations in ParC follow a phylogenetic pattern, appearing in blocks with sister taxa (Figure)
597 Interestingly, all Cluster γ isolates displayed the S80R mutation. As these point mutations are
598 at the same site, it can be reasonably assumed that this point mutation is more likely to be
599 inherited from the CC17 common ancestor than *via* several individual convergent events.
600 Thus, ciprofloxacin resistance was due to the mutations at serine 83 in the GyrA protein
601 and serine 80 in the ParC protein in all isolates sequenced for this study.
602
603 Metal resistance
604
605 No metal or biocide resistance genes were observed on any mobilome sequence. A
606 "core" metal-resistome of 3 single-copy genes (*chtR, chtS,* and *sodA*) was observed in all
607 chromosomes (SI Table 8; Figure 8). A further 3 single copy genes (*copA, copB,* and *copY/tcrY*)
608 observed in all chromosomes except for two isolates sequenced during this study, sample 9 and
609 sample 21 (both from ST192). One gene, *perR* was observed as a single-copy ortholog in all
610 clinical isolates (inclusive of ST192) but absent in all "Cork" samples. One gene, *dpr/dps* was
611 observed exclusively in sample 8 (unassigned ST). Finally, a set of 6 genes (*merA, merR1,*
612 *tcrA, tcrB, tcrY,* and *tcrZ*) were observed exclusively in the "Cork" samples (CC94). In
613 summation, all samples (except ST192) shared 7 genes (*chtR, chtS, copA, copB, copY/tcrY,*
614 *perR,* and *sodA*). These 7 genes confer resistance towards selenium (*chtR*), hydrogen peroxide
615 (*chtR, perA,* and *sodA*), chlorhexidine (*chtS*), copper (*copA, copB, copY/tcrY*) and silver (*copB*)
616 (36,133). The absence of *copABY* from ST192 may indicate copper and silver susceptibility in
617 these samples. The presence of *dpr/dps* in sample 8 is expected to confer iron resistance and
618 increased resistance to hydrogen peroxide (134,135). The presence of *merA* and *merR1* in the
619 CC94 samples indicates intrinsic resistance to mercury and phenylmercury acetate (an

organomercuric compound) and the presence of *tcrABYZ* suggests increased copper resistance (43,136–139).

Using default pangenomic parameters (106), a core chromosomal metal resistome (inclusive of soft-core) whereby a gene is represented in ≥95% genomes (≥75 genomes) comprise of six genes: *chtR, chtS, copA, copB, copY/tcrY* and *sodA.* When CC94 is excluded, the core metal resistome (≥64 genomes), is expanded to include *perR*.

Virulence factors

Each CC94 isolate contained a plasmid mediated virulence factor (VF), the biofilm-associated transcription factor *bopD* (SI Table 9; Figure 9). Only one other VF was observed to be plasmid mediated: the fibrinogen binding surface protein *fss3* (only observed within sample 18 (ST80)), all other observations were chromosomally mediated. All isolates were observed to possess *clpP* (a caseinolytic protease), *cpsA* (a biofilm associated undecaprenyl diphosphate synthase) and *cpsB* (a capsule associated phosphatidate cytidylyltransferase). A collagen adhesin precursor (*acm*) was observed in all isolates except sample 1 (ST1421) and the cell wall anchor protein *sgrA* was observed in every isolate except sample 26 (ST202). All isolates in CC17 (except Sample 1 (ST1421)) contained the bile salt hydrolase *bsh* and all CC17 isolates except ERR375031 (ST17) contained the endocarditis/biofilm-associated pilus *ebpC.* The collagen binding microbial surface components recognizing adhesive matrix molecule (MSCRAMM) gene *ecbA* was observed in sample 17 (ST80), ERR375043 (ST16), ERR374741 (ST17), ERR374767 (ST64), and in two ST203 samples (ERR374813 and ERR374914). The collagen adhesin protein *scm* was observed in 14 of 16 ST80 isolates, 5 of 17 ST17 isolates, 2 of 5 ST18, 7 of 9 ST78, 1 of 3 ST192 (sample 21) and in sample 26 (ST202). Finally, *fss3* was observed in 14 of 16 ST80 isolates, all 6 ST203 isolates, in both ST132

isolates, in one isolate each from ST16, ST17, ST64, ST78, ST192, ST787, ST1032, and ST1421 and one unclassified ST isolate (sample 22). Both *fss3* and *scm* were absent from sample 17 (ST80).

Using pangenomic parameters, a core virulome (inclusive of soft-core) whereby a gene is represented in ≥95% (≥75 genomes) would yield *acm, bopD, clpP, cpsA, cpsB, efaA, hasC, htpB, lap,* and *sgrA*, illustrating a key pathenogenic potential for biofilm-associated proteins, capsular polysaccharide biosynthesis, caseinolytic protease, endocarditis specific antigen and collagen adhesins (140,141). When only clinical isolates are considered, the core virulome (≥64 genomes) is expanded to include *bsh, epbA* and *epbC,* further increasing cardiac virulence and limiting the antibacterial function of bile acids (142,143).

Phages

Prophages were observed in the chromosome of 65 isolates (2.59±1.02; $1 \leq n \leq 4$) and all were observed to be non-transposable (SI Tables 17-19). The majority (196 of 202) of prophages were predicted to be Siphoviridae, three prophages were observed to be either Siphoviridae or Myoviridae in ERR374794, ERR374824, ERR375046, one prophage (from ERR374806) was observed to be Myoviridae, one prophage (from sample 25) was observed to be Inoviridae, and one prophage (from sample 20) was of unknown taxonomy. Prophages were not observed within any mobilome. One resistance gene, *ermB*, was observed in a prophage of ERR374741, however no other resistance mechanism (antibiotic or metal/biocide) and no virulence factors were observed on any other prophage. Without this *ermB* phage, ERR374741 would be susceptible to lincosamide, macrolide, and streptogramin. The prophage genome within ERR374741 was also observed to contain a probable secondary metabolism mechanism, however the product class could not be determined. One prophage within sample 13, sample

28, and ERR374909 was observed to contain a probable polyketide synthase. When these prophages were searched using Mash (SI Table 19), a set of 4 phages were mapped to 40 of 79 genomes. In the "Cork" samples, *Enterococcus* phage ΦFL2A was ubiquitously observed. A total of 17 prophages were observed within 17 BSAC genomes, *Enterococcus* phage EFP01 was observed in 11 genomes, *Listeria* phage B025 was observed in 4 genomes, and *Corynebacterium* phage BFK20 was observed in 2 genomes (Figure 14)

When isolation dates were considered for the BSAC samples, vancomycin resistance did not appear to be affected by the presence of a prophage as both VSE and VRE were represented (57). Both *Corynebacterium phage* BFK20 prophage-containing genomes were isolated in 2009, *Listeria* phage B025 prophage-containing genomes were isolated between 2002 and 2011, and *Enterococcus* phage EFP01 prophage-containing genomes were isolated between 2001 and 2011. In genomes sequenced for this study, *Corynebacterium* phage BFK20 was observed in 1 genome (sample 7), *Enterococcus* phage EFP01 was observed in two genomes (samples 13 and 28), *Enterococcus* phage ΦFL2A was observed in sample 11, *Listeria* phage B025 was observed in 3 genomes (samples 6, 12, and 19), and a duo of *Enterococcus* phage ΦFL2A and *Listeria* phage B025 was observed in 5 genomes (samples 2, 14, 22, 23, and 24).

A core phage has previously been observed in studies into *Enterococcus faecalis*, where it is implicated in virulence and genome plasticity (144), however, to our knowledge, this trend has not been observed for *E. faecium.*

Pangenomics

As mentioned above, a pangenome was constructed for this dataset each pangenomic category was populated (n_{genes}; "core pangenome" ($99\% \leq n_{samples} \leq 100\%$), "soft core

pangenome" (95% ≤ $n_{samples}$ < 99%), "shell pangenome" (15% ≤ $n_{samples}$ < 95%), and "cloud pangenome" ($n_{samples}$ < 15%)). The number of genes in the "whole genome" core and soft-core pangenomes and the "chromosomal" core and soft-core pangenomes are relatively unaffected by the mobilome (1324 and 236 genes *vs* 1327 and 229 genes respectively (SI Table 12)). However, the shell pangenome (2252 genes *vs* 1550) and the cloud pangenome (3306 genes *vs* 2762 genes) were more pronouncedly affected.

Pangenomic enrichment

The core pangenome (inclusive of all isolates) was observed to be enriched ($P_{BD} \leq$ 0.005) for major "house-keeping" and viability functions such as metabolism (*eg.* lipid metabolic process (GO:0006629), protein metabolic process (GO:0019538), carbohydrate metabolic process (GO:0005975), and RNA metabolic process (GO:0016070)), transport (*eg.* protein transport (GO:0015031) and ion transport (GO:0043167)), stress response (GO:0006950), signal transduction (GO:0050794), biosynthesis (GO:0009058), cellular regulation (GO:0065007), and localization (GO:0051179) (SI Table 13). The soft-core pangenome was observed to be enriched for localization and for transmembrane transport (GO:0022857). The shell pangenome was enriched for similar processes as the core genome, namely metabolic processes, transport, stress response, and localisation. The shell pangenome was also significantly enriched for small molecule metabolism (GO:0044281) and transposition (GO:0032196). The cloud pangenome was also enriched for metabolic processes, transport, and signal transduction. There was only one significant purification: transposition in the core genome. These results suggest a dynamic and non-specific evolution of the *E. faecium* pangenome and provides an insight into their genomic organisation and architecture.

720 Co-evolution

721

722 When vancomycin resistance genes were sampled from the 28 genome dataset, a
723 significant association ($P \leq 0.005$) was observed between *vanAHX* and each of *cadC, dinB,*
724 *gmuD, gmuR, hin, hisB, nag3, sdhB,* and *uxaC* and a significant dissociation with each of *clpC,*
725 *czcD, czrA, dkgB, dps, entP, fetA, fetB, fixK, fruA, gatA, lgt, mco, metB, opuCA, panE, phoP,*
726 *rihB,* and *uvrA* (SI Table 14). These results suggest that VR*Efm* carrying *vanA* typically display
727 an increase in carbohydrate processing (*via gmuD, gmuR, nag3, sdhB*, and *uxaC*), cadmium
728 transport (*via cadC*), and DNA repair (*via dinB*) (145–152). Conversely, isolates with a *vanA*+
729 genotype were also more likely to display genotypes for a decrease in biofilm formation (*via*
730 *clpC*), lipoprotein synthesis (*via lgt*), copper, iron, and zinc resistance (*via czrA fetAB,* and
731 *mco*), pantothenate biosynthesis (*via panE*), quaternary ammonium compound resistance (*via*
732 *opuCA*), oxidative resistance (*via fetAB* and *opuCA*), and stress response (*clpC, dps,* and *uvrA*)
733 (153–162). Comparatively, *vanBWXBYB* was observed to be significantly associated with each
734 of *agaS, bfrA, dgaR, dgoA, dgoD, epsF, fruA, gatB, gmuE, gmuR, gpr, Int-Tn, lacD2, manR,*
735 *manX, metK, mhpE, mro, mshA, noxE, sorC, ssbA,* and *treP* and dissociated with each of *aes,*
736 *araQ, cysM, dctM, dppC, dppE, oppB,* and *tuf.* These observations suggest that *vanB*+ VR*Efm*,
737 like *vanA*+ VR*Efm* possess increases in carbohydrate (specifically sugar) catabolic processing
738 genotypes (*via agaS, dgoAD, lacD2, manRX, mshA,* and *treB*), however, an increase was
739 observed in biofilm formation genotype (*via epsF*) and stress response (*via treB*) (163–169).
740 Isolates with a *vanB* phenotype were observed to be decreased in phenotypes for arabinose
741 uptake (*via araQ*), cysteine synthesis (*via cysM*), dicarboxylate transport (*via dctM*) and protein
742 transport (*via dppCE* and *oppB*) (170–174). These protein transport genes are essential for
743 sporulation in other Firmicute species, however, as *Enterococcus* are non-spore forming, their
744 exact role with relation to *vanB* has not been elucidated (174–176).

Interestingly, fructose processing genotypes (*via fruA*, *gatAB*, and *gpr*) are significantly decreased in *vanA*+ isolates while being significantly increased in *vanB*+ isolates (SI Table 16), suggesting a possible fructose metabolic niche in *vanB*+ isolates.

Rampant recombination

Whole genome alignments and subsequent recombination analyses indicate massive recombination events in across the CC17 genome during its divergence from non CC17 *E. faecium* ancestors (Figure 13; SI Figure 3). The majority of recombination appears to have occurred approximately within the first 50% of the genome.

Integron evolution

A total of 31 CC17 isolates were observed to possess an intrgron element, where three isolates (Samples 7, 15, and 22 (all Subcluster δ_3)) were observed to possess two (SI Table 20; Figure 15). Integrons followed a phylogenetic distribution in Subclusters δ_3 and δ_4, however all other clusters were observed to be pseudorandomly distributed throughout the phylogeny. Genes-of-interest (resistance or virulence genes) were only observed on integron containing contigs in Subclusters δ_3 and δ_4. All isolates with a gene-of-interest possessed *lsaA* (conferring resistance to clindamycin, quinupristin-dalfopristin, and dalfopristin). Subcluster δ_4 isolates only possessed *lsaA*. All Subcluster δ_3 isolates also possessed *efmA* (an efflux pump conferring resistance to macrolides and fluoroquinolones), *chtS* (conferring resistance to chlorhexidine) and *sodA* (conferring resistance to peroxides). A clade within Subcluster δ_3 (comprising of Samples 6, 12, 19, and 22) also all contained *efaA* (an endocarditis specific antigen) on their integron containing contigs.

Observation of an unusual contig

Five isolates (Samples 2, 14, 24, and 25 (all assigned to ST80; subcluster δ_4) and Sample 18 (ST80; Cluster β)) possessed a contig with a copy of *arnB*, a UDP-4-amino-4-deoxy-L-arabinose--oxoglutarate aminotransferase involved in Gram-negative Lipid A biosynthesis and polymyxin resistance (177–179). The contig in samples 2, 14, and 24 are identical (SI Figures $4_{(a)}$-$4_{(c)}$), however those attributed to samples 18 and 25 are distinct (SI Figures $4_{(d)}$-$4_{(e)}$). Samples 2 and 24 were isolated from the same patient and all three of these samples are phylogenetic neighbours (Figure 5). These contigs have a plethora of cell wall synthesis and lipopolysaccharide modification genes which may play a role in resistance. Of particular interest, sample 17 contained phage sequences suggesting a possible viral mediated horizontal gene transfer. Instances of chromosomal segment transfer between Gram-positive and Gram-negative species (and *vice-versa*) has previously been attributed to phage activity and plasmid integration (180–182).

Discussion

Within Ireland VRE has already spread nationally and is endemic to Irish hospitals (183). The level of VRE in blood stream infections in Ireland has remained at or above 30% since 2005. While some other EU countries have levels above 30%, they have not consistently had this level of VRE BSI over such a wide timeframe. In 2019, >25% of VR*Efm* BSI in eleven additional EU countries have been identified using the ECDC surveillance data. The spread and dominance of VR*Efm* within Irish hospital patients over such a long duration provides us with a unique setting to study the changing dynamics and evolution of VR*Efm*. A recent study

of the global dissemination of *E. faecium* identified that it has two main modes of genomic evolution: the acquisition and loss of genes, including antimicrobial resistance genes, through mobile genetic elements including plasmids, and homologous recombination of the chromosome. Within this global study 261 genomes contained the *vanA* gene. Within the hospital associated A1 clade of *E. faecium* the median number of AMR genes increased in the global study between 2001 and 2019 from 8 to 11 in 2015. Within this study for the *Efm* isolated between 2001 and 2011 the total AMR gene numbers corresponded with the presence of *vanA* rather than year. The vancomycin susceptible isolates and those with the *vanB* gene had less AMR genes than those with the *vanA* gene. *vanA*+ isolates had between 13 and 20 AMR genes, while those with *vanB* or no van gene had a total of between 5 and 12 AMR genes (except one vancomycin resistant isolate with *vanB* (n = 15) and one van susceptible isolate with no van genes (n = 16)). The timeframes for both groups were distributed across the ten years. Those with the highest AMR gene numbers both contained *vanA* and *vanB*. The number of AMR genes from the *vanA* study in 2018/2019 was between 12 and 18, which is consistent with the number from the previous study. Thus, the number of AMR genes in Ireland did not increase over time.

The aims of this study were to compare the pan-genomes, mobilomes and chromosomes of the available genome sequences of vancomycin resistant or susceptible *E. faecium* across the timeframe that VRE (analysed by the ECDC (2002 to 2019)) increased in prevalence from 11.1% to 38.4% in Ireland. To monitor and limit the spread of VR*Efm* we need to understand how it evolves and acquires vancomycin resistance, how transmission networks are operating and how VR*Efm* is developing resistance to last-line antibiotics.

The same AMR gene was not observed in both chromosome and mobilome in any instance, which indicated genomic control of genetic redundancy (184–186). Vancomycin resistance was observed to be either plasmid mediated (*via vanA*) or chromosomally mediated

(*via vanB*), and rarely a combination of *vanB* on the chromosome and plasmid mediated *vanA* vancomycin resistance was observed. Tetracycline (and probable tigecycline) resistance (*via tet*(*L*)) was differentially dispersed in clinical isolates, with *tet*(*L*) being observed in the mobilomes of ST17, ST18, and ST203 and in the chromosomes of ST80 (if present within the genome). Other tetracycline resistance genes implicated in tigecycline resistance, *tet*(*M*) and *tet*(*U*), were observed in addition to *tet*(*L*) on the chromosome of most ST80 isolates. Interestingly, in contrast to previously reported trends (187,188), increased metal resistance did not positively correlate with increased drug resistance, in fact the opposite was observed, with the "Cork" strains displaying diminished drug resistance and increased metal resistance.

Core-genome MLST corroborated phylogenetic and hierarchical clustering results, suggesting that core populations of *E. faecium* persist in Irish hospitals with frequent evolution towards pathogenicity *via* plasmid mediated transfer of *vanA* or chromosomal incorporation of *vanB* (Figures 5,12). This suggestion is corroborated by the genomic similarity between isolates in different clusters and clades (regardless of isolation year) and by the fact the VSE is observed within Cluster β and clades δ_1 and δ_4 (Figure 5).

We determined that a relatively stable "prophagome" exists between isolates (SI Table 17), however, these are not, to our knowledge, implicated in the evolution of resistance or pathogenicity, as observed with closely related species (189,190).

An interesting potential virulence mechanism difference was observed between *vanA*+ and *vanB*+ isolates, whereby *vanA*+ genotypes were unlikely to display a biofilm forming genotype (*via clpC*) and *vanB*+ isolates were likely to display an *epsF* biofilm forming genotype (SI Table 16). Both vancomycin resistance genotypes displayed increases in carbohydrate catabolism genotypes, however *vanB*+ isolates were more likely to display specific sugar processing genotypes. Interestingly, *vanA*+ genotypes were less likely to coevolve with metal resistance, oxidative stress resistance, panthenoate, or lipoprotein

845 synthesis genotypes. These results suggest that *vanB*+ isolates are more likely to have evolved
846 due to different stressors than *vanA*+ isolates. As *vanA* alignments used in this manuscript did
847 not have any point mutations (regardless of the year of isolation), it is likely that a narrow range
848 of mobile genetic elements confer resistance to commensal VSE in the immunocompromised,
849 prompting their pathogenicity.

850 The significant differences in observed gene density (higher in mobilomes), GC%
851 (lower in mobilomes), gene length (shorter in mobilomes) coupled with insignificant in coding
852 proportions between chromosomes and mobilomes suggest that mobilomes are more likely to
853 evolve either through the acquisition of statistically shorter sequences (while maintaining an
854 equivalent coding potential) or through successive gene fission and/or gene
855 subfunctionalisation events (191–194). While not statistically significant in most samples,
856 coding nucleotide density is, on average, approximately 7% less in mobilomes (SI Tables 11-
857 12), where significantly variant coding proportions are also observed ($\sigma^2_{chromosomal}$ = 1.04;
858 $\sigma^2_{mobilomal}$ = 3.62; $P = 1.99e^{-10}$ (Levene's test)). It is possible that these variances may just be
859 an artifact of the highly variant sequences distributed across plasmids when compared to more
860 stable chromosomes, these variances may also underly nucleotide deletion "clean up"
861 processes following subfunctionalisation or fission events indicating evolution towards a more
862 streamline, and therefore less bioeconomic expensive plasmids, thus promoting their
863 propagation (193,195–197).

864 The pangenome of *E. faecium* is relatively open, with core and soft-core components
865 only accounting for approximately 22% of the pangenome (when the mobilome is included)
866 and for 26.5% when the mobilome is excluded (SI Table 12). This variation, coupled with the
867 non-specific functional enrichments observed within the pangenome (SI Table 13), highlights
868 the rampant evolutionary capabilities in "fixed" genomic biomolecules. This variation may be
869 partially explained by the rampant recombination observed throughout CC17 or the diversity

of prophages observed throughout the respective genomes, however more research is needed to confirm the role of phages on VRE genome (Figures 11-13; SI Figure 3). As observed for phage data, integrons were observed throughout CC17 (SI Table 19; Figure 15). When integrons were observed in phylogenetic blocks, genes-of-interest were observed (SI Table 20; Figure 15). However, as these genes are widely observed in *E. faecium*, it is possible that these integrons are faciliting mobilisation of these genes. While recombination has been reported in *E. faecium* (198,199), the extent of recombination observed during this study has not, to our knowledge, been previously reported. This recombination may have been a driving force in the evolution of *E. faecium* CC17 towards pathogenicity (200). Despite the rampant variation observed between genomes used in this study, chromosomal size, gene content, gene density, and the phylogenetic proximity of each genome remained largely stable throughout the dataset. The relatedness of the VSE and VRE indicates that vancomycin susceptible *Efm* are transferring between patients and then acquiring plasmids to become VR*Efm* rather than or in addition to the movement of VRE between patients. Stability of *Efm* within Ireland over the 20 years indicates that the increases in VRE across patients is not due to frequent and multiple introductions of different *Efm* into the hospital but rather the maintenance of genomic-related susceptible *Efm* that acquire vancomycin resistance plasmids and the spread of a cohort of highly related VR*Efm*. Thus, the susceptible and resistant *Efm* are highly related and fixed within the hospital patients. To tackle VRE within Ireland we need to focus on reducing VSE in addition to reducing VR*Efm.* This requires analysis of the epidemiology of VSE, VR*Efm* and the plasmids containing the *vanA* gene. While this study provides some understanding of the past, we need to perform studies on the VS*Efm* and VR*Efm* within our hospitals to understand the continuing rise of VR*Efm* and the infection control required to limit the spread.

Conclusion

We have demonstrated that the genomic characteristics of the studies VS*Efm* and VR*Efm* (such as genome size, gene density, and gene count) are relatively stable across space and time despite an open and plastic pangenome. The relatedness of the VS*Efm* and VR*Efm* across time indicates that to reduce or remove VR*Efm* from Irish hospitals we must concurrently remove VS*Efm*. The problem of vancomycin resistance within these pathogens is plasmid mediated *vanA* rather than chromosomally mediated *vanB*.

Author contributions

Contribution	Author(s)
Conceptualisation	RL, BL, FW
Methodology	RL, FW
Software	RL
Validation	RL
Formal Analysis	RL
Investigation	RL, CM, RM
Resources	BL, FW
Data Curation	RL
Writing – Original Draft Preparation	RL, FW
Writing – Review and Editing	RL, FW
Visualisation	RL
Supervision	FW
Project Administration	FW
Funding	FW

Conflict-of-interest statement

The authors declare no conflicts of interest for this study

Acknowledgements

We wish to thank the staff in the microbiology laboratories at MMUH and the BSAC surveillance team for making their data available.

References

1. Zhou X, Willems RJL, Friedrich AW, Rossen JWA, Bathoorn E. Enterococcus faecium: From microbiological insights to practical recommendations for infection control and diagnostics. Antimicrob Resist Infect Control. 2020 Aug 10;9(1).
2. Howden BP, Holt KE, Lam MMC, Seemann T, Ballard S, Coombs GW, et al. Genomic insights to control the emergence of vancomycin-resistant enterococci. MBio. 2013 Aug 13;4(4).
3. Giannakopoulos X, Sakkas H, Ragos V, Tsiambas E, Bozidis P, Evangelou AM, et al. Impact of enterococcal urinary tract infections in immuno-compromised-neoplastic patients. JBUON. 2019;24(5):1768–75.
4. Chilambi GS, Nordstrom HR, Evans DR, Ferrolino JA, Hayden RT, Marón GM, et al. Evolution of vancomycin-resistant Enterococcus faecium during colonization and infection in immunocompromised pediatric patients. Proc Natl Acad Sci [Internet]. 2020 May 26 [cited 2021 Aug 17];117(21):11703–14. Available from: https://www.pnas.org/content/117/21/11703
5. Alghamdi F, Shakir M. The Influence of Enterococcus faecalis as a Dental Root Canal Pathogen on Endodontic Treatment: A Systematic Review. Cureus [Internet]. 2020 Mar 13 [cited 2021 Aug 17];12(3). Available from: /pmc/articles/PMC7152576/

6. Higuita NIA, Huycke MM. Enterococcal Disease, Epidemiology, and Implications for Treatment. Enterococci From Commensals to Lead Causes Drug Resist Infect [Internet]. 2014 Feb 4 [cited 2021 Aug 17]; Available from: https://www.ncbi.nlm.nih.gov/books/NBK190429/

7. Tortosa JA, Hernández-Palazón J. Enterococcus faecalis Meningitis after Spinal Anesthesia. Anesthesiology [Internet]. 2000 Mar 1 [cited 2021 Aug 17];92(3):909–909. Available from: http://pubs.asahq.org/anesthesiology/article-pdf/92/3/909/399075/0000542-200003000-00060.pdf

8. Scapellato PG, Ormazabal C, Scapellato JL, Bottaro EG, Rodríguez Brieschke MT, Vidal GI. Meningitis due to vancomycin-resistant Enterococcus faecium successfully treated with combined intravenous and intraventricular chloramphenicol [3]. J Clin Microbiol [Internet]. 2005 Jul [cited 2021 Aug 17];43(7):3578–9. Available from: https://journals.asm.org/journal/jcm

9. Tyne D Van, Gilmore MS. Friend Turned Foe: Evolution of Enterococcal Virulence and Antibiotic Resistance. Annu Rev Microbiol [Internet]. 2014 Sep 8 [cited 2021 Aug 17];68:337. Available from: /pmc/articles/PMC4384341/

10. Tsoulas C, Nathwani D. Review of meta-analyses of vancomycin compared with new treatments for Gram-positive skin and soft-tissue infections: Are we any clearer? Int J Antimicrob Agents. 2015;46(1):1–7.

11. Liu C, Bayer A, Cosgrove SE, Daum RS, Fridkin SK, Gorwitz RJ, et al. Clinical Practice Guidelines by the Infectious Diseases Society of America for the Treatment of Methicillin-Resistant Staphylococcus aureus Infections in Adults and Children: Executive Summary. Clin Infect Dis [Internet]. 2011 Feb 1 [cited 2021 Aug 17];52(3):285–92. Available from: https://academic.oup.com/cid/article/52/3/285/308819

12. Munita JM, Arias CA, Murray BE. Enterococcal Endocarditis: Can We Win the War? Curr Infect Dis Rep [Internet]. 2012 Aug [cited 2021 Aug 17];14(4):339. Available from: /pmc/articles/PMC4433165/

13. Top J, Sinnige JC, Brouwer EC, Werner G, Corander J, Severin JA, et al. Identification of a novel genomic island associated with vand-type vancomycin resistance in six Dutch vancomycin-resistant enterococcus faecium isolates. Antimicrob Agents Chemother. 2018 Mar 1;62(3).

14. Hayakawa K, Marchaim D, Palla M, Gudur UM, Pulluru H, Bathina P, et al. Epidemiology of Vancomycin-Resistant Enterococcus faecalis: a Case-Case-Control Study. Antimicrob Agents Chemother [Internet]. 2013 Jan [cited 2021 Aug 17];57(1):49. Available from: /pmc/articles/PMC3535915/

15. Machado H, Seif Y, Sakoulas G, Olson CA, Hefner Y, Anand A, et al. Environmental conditions dictate differential evolution of vancomycin resistance in Staphylococcus aureus. Commun Biol 2021 41 [Internet]. 2021 Jun 25 [cited 2021 Aug 17];4(1):1–12. Available from: https://www.nature.com/articles/s42003-021-02339-z

16. Patel R. Clinical impact of vancomycin-resistant enterococci. J Antimicrob Chemother [Internet]. 2003 [cited 2021 Aug 17];51:13–21. Available from: https://academic.oup.com/jac/article/51/suppl_3/iii13/2473485

17. Correa-Martinez C, Tönnies H, Froböse N, Mellmann A, Kampmeier S. Transmission of Vancomycin-Resistant Enterococci in the Hospital Setting: Uncovering the Patient-Environment Interplay. Microorganisms [Internet]. 2020 Feb 1 [cited 2021 Aug 17];8(2). Available from: https://pubmed.ncbi.nlm.nih.gov/32024001/

18. Courvalin P. Vancomycin Resistance in Gram-Positive Cocci. Clin Infect Dis [Internet]. 2006 Jan 1 [cited 2021 Aug 18];42(Supplement_1):S25–34. Available from: https://academic.oup.com/cid/article/42/Supplement_1/S25/275393

985 19. Arias CA, Contreras GA, Murray BE. Management of multidrug-resistant
986 enterococcalinfections. Clin Microbiol Infect [Internet]. 2010 [cited 2021 Aug
987 18];16(6):555–62. Available from: /pmc/articles/PMC3686902/
988 20. O'Driscoll T, Crank CW. Vancomycin-resistant enterococcal infections:
989 epidemiology, clinical manifestations, and optimal management. Infect Drug Resist
990 [Internet]. 2015 Jul 24 [cited 2021 Aug 17];8:217. Available from:
991 /pmc/articles/PMC4521680/
992 21. Leavis HL, Bonten MJ, Willems RJ. Identification of high-risk enterococcal clonal
993 complexes: global dispersion and antibiotic resistance. Curr Opin Microbiol. 2006
994 Oct;9(5):454–60.
995 22. Getachew Y, Hassan L, Zakaria Z, Abdul Aziz S. Genetic variability of vancomycin-
996 resistant enterococcus faecium and enterococcus faecalis isolates from humans,
997 chickens, and pigs in malaysia. Appl Environ Microbiol [Internet]. 2013 Aug [cited
998 2021 Aug 18];79(15):4528–33. Available from:
999 http://www.fas.usda.gov/dlp/circular/2011/livestock_poultry
1000 23. Leclercq R, Derlot E, Duval J, Courvalin P. Plasmid-Mediated Resistance to
1001 Vancomycin and Teicoplanin in Enterococcus Faecium. N Engl J Med. 1988 Jul
1002 21;319(3):157–61.
1003 24. Yim J, Smith JR, Rybak MJ. Role of Combination Antimicrobial Therapy for
1004 Vancomycin-Resistant Enterococcus faecium Infections: Review of the Current
1005 Evidence. Pharmacother J Hum Pharmacol Drug Ther [Internet]. 2017 May 1 [cited
1006 2021 Aug 17];37(5):579–92. Available from:
1007 https://accpjournals.onlinelibrary.wiley.com/doi/full/10.1002/phar.1922
1008 25. Zhong Z, Zhang W, Song Y, Liu W, Xu H, Xi X, et al. Comparative genomic analysis
1009 of the genus Enterococcus. Microbiol Res. 2017 Mar 1;196:95–105.

1010 26. Zhong Z, Kwok L-Y, Hou Q, Sun Y, Li W, Zhang H, et al. Comparative genomic
1011 analysis revealed great plasticity and environmental adaptation of the genomes of
1012 Enterococcus faecium. BMC Genomics [Internet]. 2019 Jul 22 [cited 2021 Aug
1013 18];20(1). Available from: /pmc/articles/PMC6647102/
1014 27. Sadowy E, Gawryszewska I, Kuch A, Żabicka D, Hryniewicz W. The changing
1015 epidemiology of VanB Enterococcus faecium in Poland. Eur J Clin Microbiol Infect
1016 Dis 2018 375 [Internet]. 2018 Feb 13 [cited 2021 Aug 24];37(5):927–36. Available
1017 from: https://link.springer.com/article/10.1007/s10096-018-3209-7
1018 28. Paoletti C, Foglia G, Princivalli MS, Magi G, Guaglianone E, Donelli G, et al. Co-
1019 transfer of vanA and aggregation substance genes from Enterococcus faecalis isolates
1020 in intra- and interspecies matings. J Antimicrob Chemother. 2007 May;59(5):1005–9.
1021 29. Handwerger S, Pucci MJ, Volk KJ, Liu J, Lee MS. The cytoplasmic peptidoglycan
1022 precursor of vancomycin-resistant Enterococcus faecalis terminates in lactate. J
1023 Bacteriol [Internet]. 1992 [cited 2021 Aug 18];174(18):5982–4. Available from:
1024 https://journals.asm.org/journal/jb
1025 30. Dutka-Malen S, Mohnas C, Arthur M, Courvalin P. The VANA glycopeptide
1026 resistance protein is related to d-alanyl-d-alanine ligase cell wall biosynthesis
1027 enzymes. Mol Gen Genet MGG 1990 2243 [Internet]. 1990 Dec [cited 2021 Aug
1028 18];224(3):364–72. Available from:
1029 https://link.springer.com/article/10.1007/BF00262430
1030 31. Roper DI, Huyton T, Vagin A, Dodson G. The molecular basis of vancomycin
1031 resistance in clinically relevant Enterococci: Crystal structure of d-alanyl-d-lactate
1032 ligase (VanA). Proc Natl Acad Sci [Internet]. 2000 Aug 1 [cited 2021 Aug
1033 18];97(16):8921–5. Available from: https://www.pnas.org/content/97/16/8921
1034 32. Clark N, Cooksey R, Hill B, Swenson J, Tenover F. Characterization of glycopeptide-

resistant enterococci from U.S. hospitals. Antimicrob Agents Chemother [Internet]. 1993 [cited 2021 Aug 18];37(11):2311–7. Available from: https://pubmed.ncbi.nlm.nih.gov/8285611/

33. Arthur M, Courvalin P. Genetics and mechanisms of glycopeptide resistance in enterococci. Antimicrob Agents Chemother [Internet]. 1993 [cited 2021 Aug 18];37(8):1563–71. Available from: https://pubmed.ncbi.nlm.nih.gov/8215264/

34. Il JP, Wee GL, Jong HS, Kyung WL, Gun JW. VanB phenotype-vanA genotype Enterococcus faecium with heterogeneous expression of teicoplanin resistance. J Clin Microbiol [Internet]. 2008 Sep [cited 2021 Aug 18];46(9):3091–3. Available from: https://journals.asm.org/journal/jcm

35. Turner RJ, Huang L-N, Viti C, Mengoni A. Metal-Resistance in Bacteria: Why Care? Genes (Basel) [Internet]. 2020 Dec 1 [cited 2021 Aug 18];11(12):1–4. Available from: /pmc/articles/PMC7764034/

36. Chen J, Li J, Zhang H, Shi W, Liu Y. Bacterial Heavy-Metal and Antibiotic Resistance Genes in a Copper Tailing Dam Area in Northern China. Front Microbiol. 2019;0(AUG):1916.

37. Li L-G, Xia Y, Zhang T. Co-occurrence of antibiotic and metal resistance genes revealed in complete genome collection. ISME J 2017 113 [Internet]. 2016 Dec 13 [cited 2021 Aug 18];11(3):651–62. Available from: https://www.nature.com/articles/ismej2016155

38. Liu H, Li M, Dao TD, Liu Y, Zhou W, Liu L, et al. Design of PdAu alloy plasmonic nanoparticles for improved catalytic performance in CO2 reduction with visible light irradiation. Vol. 26, Nano Energy. 2016.

39. Maillard J-Y. Antimicrobial biocides in the healthcare environment: efficacy, usage, policies, and perceived problems. Ther Clin Risk Manag [Internet]. 2005 Dec [cited

2021 Aug 18];1(4):307. Available from: /pmc/articles/PMC1661639/

40. Wilkinson L, White R, Chipman J. Silver and nanoparticles of silver in wound dressings: a review of efficacy and safety. J Wound Care [Internet]. 2011 [cited 2021 Aug 18];20(11):543–9. Available from: https://pubmed.ncbi.nlm.nih.gov/22240850/

41. Grass G, Rensing C, Solioz M. Metallic Copper as an Antimicrobial Surface. Appl Environ Microbiol [Internet]. 2011 Mar [cited 2021 Aug 18];77(5):1541. Available from: /pmc/articles/PMC3067274/

42. Lebreton F, Schaik W van, Sanguinetti M, Posteraro B, Torelli R, Bras F Le, et al. AsrR Is an Oxidative Stress Sensing Regulator Modulating Enterococcus faecium Opportunistic Traits, Antimicrobial Resistance, and Pathogenicity. PLOS Pathog [Internet]. 2012 Aug [cited 2021 Aug 18];8(8):e1002834. Available from: https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1002834

43. Hasman H, Aarestrup FM. tcrb, a gene conferring transferable copper resistance in Enterococcus faecium: Occurrence, transferability, and linkage to macrolide and glycopeptide resistance. Antimicrob Agents Chemother [Internet]. 2002 [cited 2021 Aug 18];46(5):1410–6. Available from: https://journals.asm.org/journal/aac

44. Pidot S, Gao W, Buultjens A, Monk I, Guerillot R, Carter G, et al. Increasing tolerance of hospital Enterococcus faecium to hand-wash alcohols. Increasing Toler Hosp Enterococcus faecium to handwash alcohols. 2016;053728.

45. Comerlato CB, Resende MCC de, Caierão J, d'Azevedo PA. Presence of virulence factors in Enterococcusfaecalis and Enterococcus faecium susceptible andresistant to vancomycin. Mem Inst Oswaldo Cruz [Internet]. 2013 [cited 2021 Aug 18];108(5):590. Available from: /pmc/articles/PMC3970601/

46. Haghi F, Lohrasbi V, Zeighami H. High incidence of virulence determinants, aminoglycoside and vancomycin resistance in enterococci isolated from hospitalized

1085 patients in Northwest Iran. BMC Infect Dis 2019 191 [Internet]. 2019 Aug 27 [cited
1086 2021 Aug 18];19(1):1–10. Available from:
1087 https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-019-4395-3
1088 47. Biswas PP, Dey S, Adhikari L, Sen A. Virulence Markers of Vancomycin Resistant
1089 Enterococci Isolated from Infected and Colonized Patients. J Glob Infect Dis
1090 [Internet]. 2014 Oct 1 [cited 2021 Aug 18];6(4):157. Available from:
1091 /pmc/articles/PMC4265831/
1092 48. Kim EB, Marco ML. Nonclinical and clinical enterococcus faecium strains, but not
1093 enterococcus faecalis strains, have distinct structural and functional genomic features.
1094 Appl Environ Microbiol [Internet]. 2014 Jan [cited 2021 Aug 18];80(1):154–65.
1095 Available from: http://dx.doi.org/10.1128
1096 49. van Hal SJ, Willems RJL, Gouliouris T, Ballard SA, Coque TM, Hammerum AM, et
1097 al. The global dissemination of hospital clones of Enterococcus faecium. Genome Med
1098 2021 131 [Internet]. 2021 Mar 30 [cited 2021 Aug 18];13(1):1–12. Available from:
1099 https://genomemedicine.biomedcentral.com/articles/10.1186/s13073-021-00868-0
1100 50. Banfield And JF, Young M. Variety-the splice of life-in microbial communities.
1101 Science (80-). 2009 Nov 27;326(5957):1198–9.
1102 51. Koskella B, Brockhurst MA. Bacteria–phage coevolution as a driver of ecological and
1103 evolutionary processes in microbial communities. FEMS Microbiol Rev [Internet].
1104 2014 Sep 1 [cited 2021 Aug 18];38(5):916–31. Available from:
1105 https://academic.oup.com/femsre/article/38/5/916/497097
1106 52. Broniewski JM, Meaden S, Paterson S, Buckling A, Westra ER. The effect of phage
1107 genetic diversity on bacterial resistance evolution. ISME J 2020 143 [Internet]. 2020
1108 Jan 2 [cited 2021 Aug 18];14(3):828–36. Available from:
1109 https://www.nature.com/articles/s41396-019-0577-7

1110 53. Bonilla N, Santiago T, Marcos P, Urdaneta M, Santo Domingo J, Toranzos GA.
1111 Enterophages, a group of phages infecting Enterococcus faecalis, and their potential as
1112 alternate indicators of human faecal contamination. Water Sci Technol.
1113 2010;61(2):293–300.
1114 54. Lee D, Im J, Na H, Ryu S, Yun C-H, Han SH. The Novel Enterococcus Phage
1115 vB_EfaS_HEf13 Has Broad Lytic Activity Against Clinical Isolates of Enterococcus
1116 faecalis. Front Microbiol. 2019 Dec 17;0:2877.
1117 55. Wandro S, Oliver A, Gallagher T, Weihe C, England W, Martiny JBH, et al.
1118 Predictable Molecular Adaptation of Coevolving Enterococcus faecium and Lytic
1119 Phage EfV12-phi1. Front Microbiol. 2019;0(JAN):3192.
1120 56. Whelan F, Rusilowicz M, McInerney J. Coinfinder: Detecting significant associations
1121 and dissociations in pangenomes. Microb Genomics [Internet]. 2020 [cited 2021 Aug
1122 18];6(3). Available from: https://pubmed.ncbi.nlm.nih.gov/32100706/
1123 57. Raven KE, Reuter S, Gouliouris T, Reynolds R, Russell JE, Brown NM, et al.
1124 Genome-based characterization of hospital-adapted Enterococcus faecalis lineages.
1125 Nat Microbiol 2016 13 [Internet]. 2016 Feb 8 [cited 2021 Aug 24];1(3):1–7. Available
1126 from: https://www.nature.com/articles/nmicrobiol201533
1127 58. Leinonen R, Sugawara H, Shumway M. The Sequence Read Archive. Nucleic Acids
1128 Res [Internet]. 2011 Jan [cited 2021 Aug 18];39(Database issue):D19. Available from:
1129 /pmc/articles/PMC3013647/
1130 59. Krueger F. Babraham Bioinformatics - Trim Galore! [Internet]. 2012 [cited 2021 Jan
1131 16]. Available from: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
1132 60. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing
1133 reads. EMBnet.journal [Internet]. 2011 May 2 [cited 2021 Jan 16];17(1):10. Available
1134 from: http://www-huber.embl.de/users/an-

1135 61. Andrews S, Krueger F, Seconds-Pichon A, Biggins F, Wingett S. FastQC. A quality
1136 control tool for high throughput sequence data. Babraham Bioinformatics. 2015 [cited
1137 2021 Aug 18];1(1):undefined-undefined. Available from:
1138 https://www.mendeley.com/catalogue/8057171e-e700-36a0-b936-
1139 1c307058462d/?utm_source=desktop&utm_medium=1.19.8&utm_campaign=open_ca
1140 talog&userDocumentId=%7B124edc8f-5459-3c1b-a449-8ab9bc3e95f8%7D
1141 62. Wick RR, Judd LM, Gorrie CL, Holt KE. Completing bacterial genome assemblies
1142 with multiplex MinION sequencing. Microb Genomics. 2017 Oct 1;3(10):e000132.
1143 63. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al.
1144 SPAdes: A new genome assembly algorithm and its applications to single-cell
1145 sequencing. J Comput Biol. 2012 May 1;19(5):455–77.
1146 64. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods
1147 [Internet]. 2012 [cited 2017 Sep 6];9(4):357–9. Available from:
1148 http://www.nature.com/articles/nmeth.1923
1149 65. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: An
1150 Integrated Tool for Comprehensive Microbial Variant Detection and Genome
1151 Assembly Improvement. PLoS One [Internet]. 2014 Nov 19 [cited 2021 Aug
1152 18];9(11):e112963. Available from:
1153 https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0112963
1154 66. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al.
1155 BLAST+: architecture and applications. BMC Bioinformatics [Internet]. 2009 Dec 15
1156 [cited 2017 Aug 1];10(1):421. Available from: http://www.biomedcentral.com/1471-
1157 2105/10/421
1158 67. Altschul SSF, Gish W, Miller W, Myers EEW, Lipman DJD. Basic local alignment
1159 search tool [Internet]. Journal of Molecular Biology Oct, 1990 p. 403–10. Available

from: https://linkinghub.elsevier.com/retrieve/pii/S0022283605803602

68. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics [Internet]. 2009 Aug 15 [cited 2018 Sep 12];25(16):2078–9. Available from: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btp352

69. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res [Internet]. 2015 Jul 1 [cited 2021 Aug 18];25(7):1043. Available from: /pmc/articles/PMC4484387/

70. Jolley KA, Maiden MCJ. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics [Internet]. 2010 Dec 10 [cited 2020 Apr 14];11(1):595. Available from: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-11-595

71. Seemann T. GitHub - tseemann/mlst: Scan contig files against PubMLST typing schemes [Internet]. 2014 [cited 2021 Aug 18]. Available from: https://github.com/tseemann/mlst

72. Jolley K, Bray J, Maiden M. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome open Res [Internet]. 2018 [cited 2021 Aug 26];3. Available from: https://pubmed.ncbi.nlm.nih.gov/30345391/

73. Schwengers O, Barth P, Falgenhauer L, Hain T, Chakraborty T, Goesmann A. Platon: identification and characterization of bacterial plasmid contigs in short-read draft assemblies exploiting protein sequence-based replicon distribution scores. Microb genomics [Internet]. 2020 [cited 2021 Aug 18];6(10):1–12. Available from:

https://pubmed.ncbi.nlm.nih.gov/32579097/

74. Angelopoulou A, Warda A, O'Connor P, Stockdale S, Shkoporov A, Field D, et al. Diverse Bacteriocins Produced by Strains From the Human Milk Microbiota. Front Microbiol [Internet]. 2020 May 19 [cited 2021 Aug 24];11. Available from: https://pubmed.ncbi.nlm.nih.gov/32508758/

75. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, et al. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol 2016 171 [Internet]. 2016 Jun 20 [cited 2021 Aug 18];17(1):1–14. Available from: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-0997-x

76. Galata V, Fehlmann T, Backes C, Keller A. PLSDB: a resource of complete bacterial plasmids. Nucleic Acids Res [Internet]. 2019 [cited 2021 Nov 2];47(D1):D195–202. Available from: https://pubmed.ncbi.nlm.nih.gov/30380090/

77. Seemann T. Prokka: Rapid prokaryotic genome annotation. Bioinformatics. 2014 Jul 15;30(14):2068–9.

78. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinforma 2010 111 [Internet]. 2010 Mar 8 [cited 2021 Aug 18];11(1):1–11. Available from: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-11-119

79. Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res [Internet]. 2004 [cited 2021 Aug 18];32(1):11–6. Available from: https://pubmed.ncbi.nlm.nih.gov/14704338/

80. Skennerton C. minced: Mining CRISPRs in Environmental Datasets [Internet]. 2013 [cited 2021 Aug 18]. Available from: https://github.com/ctSkennerton/minced

81. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal

peptides from transmembrane regions. 2011 Sep 29 [cited 2017 Sep 6];8(10):785–6. Available from: http://www.nature.com/doifinder/10.1038/nmeth.1701

82. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res [Internet]. 2011 Jul 1 [cited 2021 Aug 18];39(Web Server issue):W29. Available from: /pmc/articles/PMC3125773/

83. Stajich J, Block D, Boulez K, Brenner S, Chervitz S, Dagdigian C, et al. The Bioperl toolkit: Perl modules for the life sciences. Genome Res [Internet]. 2002 Oct 1 [cited 2021 Aug 18];12(10):1611–8. Available from: https://pubmed.ncbi.nlm.nih.gov/12368254/

84. Seemann T. barrnap: Bacterial ribosomal RNA predictor [Internet]. 2013 [cited 2021 Aug 18]. Available from: https://github.com/tseemann/barrnap

85. Jones P, Binns D, Chang H-Y, Fraser M, Li W, Mcanulla C, et al. InterProScan 5: genome-scale protein function classification. 2014;30(9):1236–40.

86. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, et al. The Pfam protein families database. Nucleic Acids Res [Internet]. 2004 Jan 1 [cited 2019 Sep 29];32(90001):138D – 141. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14681378

87. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet [Internet]. 2000 May [cited 2018 Sep 13];25(1):25–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10802651

88. Carroll LM, Larralde M, Fleck JS, Ponnudurai R, Milanese A, Cappio E, et al. Accurate de novo identification of biosynthetic gene clusters with GECCO. bioRxiv [Internet]. 2021 May 4 [cited 2021 Aug 25];2021.05.03.442509. Available from: https://www.biorxiv.org/content/10.1101/2021.05.03.442509v1

89. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, et al. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res [Internet]. 2019 Jul 2 [cited 2021 Aug 18];47(W1):W81–7. Available from: https://academic.oup.com/nar/article/47/W1/W81/5481154

90. Johansson MHK, Bortolaia V, Tansirichaiya S, Aarestrup FM, Roberts AP, Petersen TN. Detection of mobile genetic elements associated with antibiotic resistance in Salmonella enterica using a newly developed web tool: MobileElementFinder. J Antimicrob Chemother [Internet]. 2021 Jan 1 [cited 2021 Sep 1];76(1):101–9. Available from: https://academic.oup.com/jac/article/76/1/101/5917579

91. Seemann. T. ABRicate: mass screening of contigs for antimicrobial resistance or virulence genes. [Internet]. 2014 [cited 2021 Aug 18]. Available from: https://github.com/tseemann/abricate/commits/master?after=955d402a23371a61bfca48f1b9e0d30ac724e6ef+209&branch=master

92. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res. 2020 Jan 8;48(D1):D517–25.

93. Zankari E, Allesøe R, Joensen KG, Cavaco LM, Lund O, Aarestrup FM. PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. J Antimicrob Chemother [Internet]. 2017 Oct 1 [cited 2021 Aug 25];72(10):2764–8. Available from: https://academic.oup.com/jac/article/72/10/2764/3979530

94. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, et al. VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res [Internet]. 2005 Jan 1 [cited 2021 Aug 18];33(Database issue). Available from: https://pubmed.ncbi.nlm.nih.gov/15608208/

95. Pal C, Bengtsson-Palme J, Rensing C, Kristiansson E, Larsson DGJ. BacMet:

antibacterial biocide and metal resistance genes database [Internet]. Nucleic Acids Research Oxford University Press; Jan 1, 2014 p. D737. Available from: /pmc/articles/PMC3965030/

96. Rice P, Longden L, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet [Internet]. 2000 Jun 1 [cited 2021 Nov 22];16(6):276–7. Available from: https://pubmed.ncbi.nlm.nih.gov/10827456/

97. Welch BL. The generalization of "Student's" problem with several different population varlances are involved. Biometrika [Internet]. 1947 Jan 1 [cited 2021 Jan 16];34(1–2):28–35. Available from: https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/34.1-2.28

98. Student. The Probable Error of a Mean. Biometrika. 1908 Mar;6(1):1.

99. Bonferroni C. Teoria statistica delle classi e calcolo delle probabilita. Pubbl del R Ist Super di Sci Econ e Commericiali di Firenze [Internet]. 1936 [cited 2021 Apr 20];8:3–62. Available from: https://ci.nii.ac.jp/naid/20001561442

100. Dunn OJ. Multiple Comparisons among Means. J Am Stat Assoc [Internet]. 1961 Mar [cited 2019 Feb 12];56(293):52–64. Available from: http://www.tandfonline.com/doi/abs/10.1080/01621459.1961.10482090

101. Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers EJ, Berk R, et al. Redefine statistical significance [Internet]. Vol. 2, Nature Human Behaviour. Nature Publishing Group; 2018 [cited 2021 Apr 27]. p. 6–10. Available from: www.nature.com/nathumbehav

102. Kolmogorov AN. Sulla determinazione empirica di una lgge di distribuzione – ScienceOpen. Inst Ital Attuari, Giorn [Internet]. 1933 [cited 2021 May 5]; Available from: https://www.scienceopen.com/document?vid=c3c08573-63b2-4153-a72e-97bd1b3663a0

1285 103. Lilliefors HW. On the Kolmogorov-Smirnov Test for Normality with Mean and
1286 Variance Unknown. J Am Stat Assoc. 1967;62(318):399–402.
1287 104. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and
1288 manipulating networks. In: International AAAI Conference on Weblogs and Social
1289 Media. 2009.
1290 105. Fruchterman TMJ, Reingold EM. Graph drawing by force-directed placement. Softw
1291 Pract Exp [Internet]. 1991 Nov 1 [cited 2021 Nov 10];21(11):1129–64. Available
1292 from: https://onlinelibrary.wiley.com/doi/full/10.1002/spe.4380211102
1293 106. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary:
1294 Rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015 May
1295 25;31(22):3691–3.
1296 107. Löytynoja A. Phylogeny-aware alignment with PRANK. Methods Mol Biol [Internet].
1297 2014 [cited 2020 Aug 10];1079:155–70. Available from:
1298 https://pubmed.ncbi.nlm.nih.gov/24170401/
1299 108. Fisher RA. On the “Probable Error” of a Coefficient of Correlation Deduced from a
1300 Small Sample. Metron. 1921;1:3–32.
1301 109. Klopfenstein D V., Zhang L, Pedersen BS, Ramírez F, Warwick Vesztrocy A, Naldi
1302 A, et al. GOATOOLS: A Python library for Gene Ontology analyses. Sci Rep
1303 [Internet]. 2018 Dec 18 [cited 2018 Dec 16];8(1):10872. Available from:
1304 http://www.nature.com/articles/s41598-018-28948-z
1305 110. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated
1306 alignment trimming in large-scale phylogenetic analyses. Bioinformatics [Internet].
1307 2009 Aug 1 [cited 2018 Sep 12];25(15):1972–3. Available from:
1308 https://academic.oup.com/bioinformatics/article-
1309 lookup/doi/10.1093/bioinformatics/btp348

111. Kück P, Meusemann K. FASconCAT: Convenient handling of data matrices. Mol Phylogenet Evol [Internet]. 2010 Sep [cited 2017 Aug 1];56(3):1115–8. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1055790310001909

112. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol [Internet]. 2015 Jan [cited 2018 Sep 11];32(1):268–74. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25371430

113. Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. Mol Biol Evol [Internet]. 2020 Jan 1 [cited 2021 Aug 18];37(1):291–4. Available from: https://academic.oup.com/mbe/article/37/1/291/5552155

114. Tavare S. Some probabilistic and statistical problems in the analysis of DNA sequences. Lect Math life [Internet]. 1986 [cited 2021 Aug 18]; Available from: https://books.google.com/books?hl=en&lr=&id=8aI1phhOKhgC&oi=fnd&pg=PA57&ots=roMG4RHdKi&sig=csERmeVhWARlEZ2mHiyf3_zHGSA

115. Lechner M, Findeiß S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: Detection of (Co-)orthologs in large-scale analysis. BMC Bioinformatics [Internet]. 2011 Apr 28 [cited 2020 Jun 30];12(1):124. Available from: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-124

116. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res [Internet]. 2004 Mar 8 [cited 2018 Sep 13];32(5):1792–7. Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkh340

117. Le SQ, Gascuel O. An Improved General Amino Acid Replacement Matrix. Mol Biol Evol [Internet]. 2008 Apr 3 [cited 2018 Jan 5];25(7):1307–20. Available from:

http://www.ncbi.nlm.nih.gov/pubmed/18367465

118. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res [Internet]. 2021 Jul 2 [cited 2021 Aug 27];49(W1):W293–6. Available from: https://academic.oup.com/nar/article/49/W1/W293/6246398

119. Cheng L, Connor T, Sirén J, Aanensen D, Corander J. Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. Mol Biol Evol [Internet]. 2013 May [cited 2021 Aug 31];30(5):1224–8. Available from: https://pubmed.ncbi.nlm.nih.gov/23408797/

120. Tonkin-Hill G, Lees JA, Bentley SD, Frost SDW, Corander J. RhierBAPS: An R implementation of the population clustering algorithm hierBAPS. Wellcome Open Res [Internet]. 2018 [cited 2021 Aug 31];3. Available from: /pmc/articles/PMC6178908/

121. Silva M, Machado MP, Silva DN, Rossi M, Moran-Gilad J, Santos S, et al. chewBBACA: A complete suite for gene-by-gene schema creation and strain identification. Microb Genomics [Internet]. 2018 Mar 1 [cited 2021 Aug 27];4(3). Available from: /pmc/articles/PMC5885018/

122. Zhou Z, Alikhan N-F, Sergeant MJ, Luhmann N, Vaz C, Francisco AP, et al. GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens. Genome Res [Internet]. 2018 Sep 1 [cited 2021 Aug 27];28(9):1395–404. Available from: https://genome.cshlp.org/content/28/9/1395.full

123. Abudahab K, Prada JM, Yang Z, Bentley SD, Croucher NJ, Corander J, et al. PANINI: Pangenome Neighbour Identification for Bacterial Populations. Microb Genomics [Internet]. 2018 Nov 22 [cited 2021 Aug 31];5(4):e000220. Available from: https://www.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.000220

124. Argimón S, Abudahab K, Goater RJE, Fedosejev A, Bhai J, Glasner C, et al.

Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. Microb Genomics [Internet]. 2016 Nov 30 [cited 2021 Sep 1];2(11):e000093. Available from: https://www.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.000093

125. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res [Internet]. 2015 Feb 18 [cited 2021 Sep 1];43(3):e15–e15. Available from: https://academic.oup.com/nar/article/43/3/e15/2410982

126. Hadfield J, Croucher NJ, Goater RJ, Abudahab K, Aanensen DM, Harris SR. Phandango: an interactive viewer for bacterial population genomics. Bioinformatics [Internet]. 2018 Jan 15 [cited 2021 Sep 1];34(2):292–3. Available from: https://academic.oup.com/bioinformatics/article/34/2/292/4212949

127. Starikova E V, Tikhonova PO, Prianichnikov NA, Rands CM, Zdobnov EM, Ilina EN, et al. Phigaro: high-throughput prophage sequence annotation. Bioinformatics [Internet]. 2020 Jun 1 [cited 2021 Aug 18];36(12):3882–4. Available from: https://academic.oup.com/bioinformatics/article/36/12/3882/5822875

128. Cury J, Jové T, Touchon M, Néron B, Rocha EP. Identification and analysis of integrons and cassette arrays in bacterial genomes. Nucleic Acids Res [Internet]. 2016 Jun 2 [cited 2021 Nov 22];44(10):4539–50. Available from: https://pubmed.ncbi.nlm.nih.gov/27130947/

129. Singh K, Malathum K, Murray B. Disruption of an Enterococcus faecium species-specific gene, a homologue of acquired macrolide resistance genes of staphylococci, is associated with an increase in macrolide susceptibility. Antimicrob Agents Chemother [Internet]. 2001 [cited 2021 Aug 24];45(1):263–6. Available from:

https://pubmed.ncbi.nlm.nih.gov/11120975/

130. Costa Y, Galimand M, Leclercq R, Duval J, Courvalin P. Characterization of the chromosomal aac(6')-Ii gene specific for Enterococcus faecium. Antimicrob Agents Chemother [Internet]. 1993 [cited 2021 Aug 24];37(9):1896–903. Available from: https://pubmed.ncbi.nlm.nih.gov/8239603/

131. Draker K, Northrop D, Wright G. Kinetic mechanism of the GCN5-related chromosomal aminoglycoside acetyltransferase AAC(6')-Ii from Enterococcus faecium: evidence of dimer subunit cooperativity. Biochemistry [Internet]. 2003 Jun 3 [cited 2021 Aug 24];42(21):6565–74. Available from: https://pubmed.ncbi.nlm.nih.gov/12767240/

132. Wybenga-Groot L, Draker K, Wright G, Berghuis A. Crystal structure of an aminoglycoside 6'-N-acetyltransferase: defining the GCN5-related N-acetyltransferase superfamily fold. Structure [Internet]. 1999 [cited 2021 Aug 24];7(5):497–507. Available from: https://pubmed.ncbi.nlm.nih.gov/10378269/

133. Prieto AMG, Wijngaarden J, Braat JC, Rogers MRC, Majoor E, Brouwer EC, et al. The Two-Component System ChtRS Contributes to Chlorhexidine Tolerance in Enterococcus faecium. Antimicrob Agents Chemother [Internet]. 2017 May 1 [cited 2021 Aug 24];61(5). Available from: /pmc/articles/PMC5404517/

134. Fujishima K, Kawada-Matsuo M, Oogai Y, Tokuda M, Torii M, Komatsuzawa H. dpr and sod in Streptococcus mutans Are Involved in Coexistence with S. sanguinis, and PerR Is Associated with Resistance to H2O2. Appl Environ Microbiol [Internet]. 2013 Mar [cited 2021 Aug 24];79(5):1436. Available from: /pmc/articles/PMC3591952/

135. Yamamoto Y, Poole LB, Hantgan RR, Kamio Y. An iron-binding protein, Dpr, from Streptococcus mutans prevents iron-dependent hydroxyl radical formation in vitro. J Bacteriol [Internet]. 2002 [cited 2021 Aug 24];184(11):2931–9. Available from:

http://www.jphilo

136. Hasman H, Kempf I, Chidaine B, Cariolet R, Ersbøll AK, Houe H, et al. Copper resistance in Enterococcus faecium, mediated by the tcrB gene, is selected by supplementation of pig feed with copper sulfate. Appl Environ Microbiol [Internet]. 2006 Sep [cited 2021 Aug 24];72(9):5784–9. Available from: https://journals.asm.org/journal/aem

137. Mazaheri Nezhad Fard R, Heuzenroeder MW, Barton MD. Antimicrobial and heavy metal resistance in commensal enterococci isolated from pigs. Vet Microbiol. 2011 Mar 24;148(2–4):276–82.

138. Horsburgh MJ, Clements MO, Crossley H, Ingham E, Foster SJ. PerR Controls Oxidative Stress Resistance and Iron Storage Proteins and Is Required for Virulence in Staphylococcus aureus. Infect Immun [Internet]. 2001 [cited 2021 Aug 24];69(6):3744. Available from: /pmc/articles/PMC98383/

139. Møller AK, Barkay T, Hansen MA, Norman A, Hansen LH, Sørensen SJ, et al. Mercuric reductase genes (merA) and mercury resistance plasmids in High Arctic snow, freshwater and sea-ice brine. FEMS Microbiol Ecol [Internet]. 2014 Jan 1 [cited 2021 Aug 24];87(1):52–63. Available from: https://academic.oup.com/femsec/article/87/1/52/508980

140. Zaheer R, Cook SR, Barbieri R, Goji N, Cameron A, Petkau A, et al. Surveillance of Enterococcus spp. reveals distinct species and antimicrobial resistance diversity across a One-Health continuum. Sci Reports 2020 101 [Internet]. 2020 Mar 3 [cited 2021 Aug 25];10(1):1–16. Available from: https://www.nature.com/articles/s41598-020-61002-5

141. Creti R, Koch S, Fabretti F, Baldassarri L, Huebner J. Enterococcal colonization of the gastro-intestinal tract: role of biofilm and environmental oligosaccharides. BMC

1435 Microbiol [Internet]. 2006 Jul 11 [cited 2021 Aug 25];6:60. Available from:
1436 /pmc/articles/PMC1534043/
1437 142. Nallapareddy SR, Singh K V., Sillanpää J, Garsin DA, Höök M, Erlandsen SL, et al.
1438 Endocarditis and biofilm-associated pili of Enterococcus faecalis. J Clin Invest
1439 [Internet]. 2006 Oct 2 [cited 2021 Aug 25];116(10):2799. Available from:
1440 /pmc/articles/PMC1578622/
1441 143. Dahl A, Rasmussen R V., Bundgaard H, Hassager C, Bruun LE, Lauridsen TK, et al.
1442 Enterococcus faecalis Infective Endocarditis. Circulation [Internet]. 2013 Apr 30 [cited
1443 2021 Aug 17];127(17):1810–7. Available from:
1444 https://www.ahajournals.org/doi/abs/10.1161/circulationaha.112.001170
1445 144. Matos RC, Lapaque N, Rigottier-Gois L, Debarbieux L, Meylheuc T, Gonzalez-Zorn
1446 B, et al. Enterococcus faecalis Prophage Dynamics and Contributions to Pathogenic
1447 Traits. PLoS Genet [Internet]. 2013 Jun [cited 2021 Aug 25];9(6). Available from:
1448 /pmc/articles/PMC3675006/
1449 145. Mayer C, Vocadlo D, Mah M, Rupitz K, Stoll D, Warren R, et al. Characterization of a
1450 beta-N-acetylhexosaminidase and a beta-N-acetylglucosaminidase/beta-glucosidase
1451 from Cellulomonas fimi. FEBS J [Internet]. 2006 Jul [cited 2021 Sep
1452 2];273(13):2929–41. Available from: https://pubmed.ncbi.nlm.nih.gov/16762038/
1453 146. Rivolta C, Soldo B, Lazarevic V, Joris B, Mauël C, Karamat D. A 35.7 kb DNA
1454 fragment from the Bacillus subtilis chromosome containing a putative 12.3 kb operon
1455 involved in hexuronate catabolism and a perfectly symmetrical hypothetical catabolite-
1456 responsive element. Microbiology [Internet]. 1998 [cited 2021 Sep 2];144 (Pt
1457 4(4):877–84. Available from: https://pubmed.ncbi.nlm.nih.gov/9579062/
1458 147. Phillips M, Hederstedt L, Hasnain S, Rutberg L, Guest J. Nucleotide sequence
1459 encoding the flavoprotein and iron-sulfur protein subunits of the Bacillus subtilis PY79

succinate dehydrogenase complex. J Bacteriol [Internet]. 1987 [cited 2021 Sep 2];169(2):864–73. Available from: https://pubmed.ncbi.nlm.nih.gov/3027051/

148. Wipat A, Carter N, Brignell S, Guy B, Piper K, Sanders J, et al. The dnaB-pheA (256 degrees-240 degrees) region of the Bacillus subtilis chromosome containing genes responsible for stress responses, the utilization of plant cell walls and primary metabolism. Microbiology [Internet]. 1996 [cited 2021 Sep 2];142 (Pt 1(11):3067–78. Available from: https://pubmed.ncbi.nlm.nih.gov/8969504/

149. Mekjian K, Bryan E, Beall B, Moran C. Regulation of hexuronate utilization in Bacillus subtilis. J Bacteriol [Internet]. 1999 [cited 2021 Sep 2];181(2):426–33. Available from: https://pubmed.ncbi.nlm.nih.gov/9882655/

150. Setlow B, Cabrera-Hernandez A, Cabrera-Martinez R, Setlow P. Identification of aryl-phospho-beta-D-glucosidases in Bacillus subtilis. Arch Microbiol [Internet]. 2004 Jan [cited 2021 Sep 2];181(1):60–7. Available from: https://pubmed.ncbi.nlm.nih.gov/14652714/

151. Sadaie Y, Nakadate H, Fukui R, Yee L, Asai K. Glucomannan utilization operon of Bacillus subtilis. FEMS Microbiol Lett [Internet]. 2008 Feb [cited 2021 Sep 2];279(1):103–9. Available from: https://pubmed.ncbi.nlm.nih.gov/18177310/

152. Wong M, Lin Y, Rosen B. The soft metal ion binding sites in the Staphylococcus aureus pI258 CadC Cd(II)/Pb(II)/Zn(II)-responsive repressor are formed between subunits of the homodimer. J Biol Chem [Internet]. 2002 Oct 25 [cited 2021 Sep 2];277(43):40930–6. Available from: https://pubmed.ncbi.nlm.nih.gov/12176999/

153. Msadek T, Kunst F, Rapoport G. MecB of Bacillus subtilis, a member of the ClpC ATPase family, is a pleiotropic regulator controlling competence gene expression and growth at high temperature. Proc Natl Acad Sci U S A [Internet]. 1994 Jun 1 [cited 2021 Sep 2];91(13):5788–92. Available from:

https://europepmc.org/articles/PMC44082

154. Krüger E, Völker U, Hecker M. Stress induction of clpC in Bacillus subtilis and its involvement in stress tolerance. J Bacteriol [Internet]. 1994 [cited 2021 Sep 2];176(11):3360–7. Available from: https://pubmed.ncbi.nlm.nih.gov/8195092/

155. Leskelä S, Wahlström E, Kontinen V, Sarvas M. Lipid modification of prelipoproteins is dispensable for growth but essential for efficient protein secretion in Bacillus subtilis: characterization of the Lgt gene. Mol Microbiol [Internet]. 1999 [cited 2021 Sep 2];31(4):1075–85. Available from: https://pubmed.ncbi.nlm.nih.gov/10096076/

156. Harvie D, Andreini C, Cavallaro G, Meng W, Connolly B, Yoshida K, et al. Predicting metals sensed by ArsR-SmtB repressors: allosteric interference by a non-effector metal. Mol Microbiol [Internet]. 2006 Feb [cited 2021 Sep 2];59(4):1341–56. Available from: https://pubmed.ncbi.nlm.nih.gov/16430705/

157. Nicolaou S, Fast A, Nakamaru-Ogiso E, Papoutsakis E. Overexpression of fetA (ybbL) and fetB (ybbM), Encoding an Iron Exporter, Enhances Resistance to Oxidative Stress in Escherichia coli. Appl Environ Microbiol [Internet]. 2013 Dec [cited 2021 Sep 2];79(23):7210–9. Available from: https://pubmed.ncbi.nlm.nih.gov/24038693/

158. Zheng R, Blanchard J. Kinetic and mechanistic analysis of the E. coli panE-encoded ketopantoate reductase. Biochemistry [Internet]. 2000 Apr 4 [cited 2021 Sep 2];39(13):3708–17. Available from: https://pubmed.ncbi.nlm.nih.gov/10736170/

159. Fraser K, Harvie D, Coote P, O'Byrne C. Identification and characterization of an ATP binding cassette L-carnitine transporter in Listeria monocytogenes. Appl Environ Microbiol [Internet]. 2000 [cited 2021 Sep 2];66(11):4696–704. Available from: https://pubmed.ncbi.nlm.nih.gov/11055912/

160. Huynh T, Choi P, Sureka K, Ledvina H, Campillo J, Tong L, et al. Cyclic di-AMP targets the cystathionine beta-synthase domain of the osmolyte transporter OpuC. Mol

Microbiol [Internet]. 2016 Oct 1 [cited 2021 Sep 2];102(2):233–43. Available from: https://pubmed.ncbi.nlm.nih.gov/27378384/

161. Pruteanu M, Baker T. Controlled degradation by ClpXP protease tunes the levels of the excision repair protein UvrA to the extent of DNA damage. Mol Microbiol [Internet]. 2009 Feb [cited 2021 Sep 2];71(4):912–24. Available from: https://pubmed.ncbi.nlm.nih.gov/19183285/

162. Myles G, Sancar A. Isolation and characterization of functional domains of UvrA. Biochemistry [Internet]. 1991 Apr 1 [cited 2021 Sep 2];30(16):3834–40. Available from: https://pubmed.ncbi.nlm.nih.gov/1826851/

163. Cooper R. The utilisation of D-galactonate and D-2-oxo-3-deoxygalactonate by Escherichia coli K-12. Biochemical and genetical studies. Arch Microbiol [Internet]. 1978 Aug [cited 2021 Sep 2];118(2):199–206. Available from: https://pubmed.ncbi.nlm.nih.gov/211976/

164. LowKam C, Liotard B, Sygusch J. Structure of a class I tagatose-1,6-bisphosphate aldolase: investigation into an apparent loss of stereospecificity. J Biol Chem [Internet]. 2010 Jul 2 [cited 2021 Sep 2];285(27):21143–52. Available from: https://pubmed.ncbi.nlm.nih.gov/20427286/

165. Erni B, Zanolari B, Kocher H. The mannose permease of Escherichia coli consists of three different proteins. Amino acid sequence and function in sugar transport, sugar phosphorylation, and penetration of phage lambda DNA. J Biol Chem [Internet]. 1987 Apr 1 [cited 2021 Sep 2];262(11):5238–47. Available from: https://europepmc.org/article/MED/2951378

166. Stolz B, Huber M, Marković-Housley Z, Erni B. The mannose transporter of Escherichia coli. Structure and function of the IIABMan subunit. J Biol Chem [Internet]. 1993 Dec 1 [cited 2021 Sep 2];268(36):27094–9. Available from:

https://europepmc.org/article/MED/8262947

167. Newton G, Koledin T, Gorovitz B, Rawat M, Fahey R, Av-Gay Y. The glycosyltransferase gene encoding the enzyme catalyzing the first step of mycothiol biosynthesis (mshA). J Bacteriol [Internet]. 2003 Jun [cited 2021 Sep 2];185(11):3476–9. Available from: https://pubmed.ncbi.nlm.nih.gov/12754249/

168. Schöck F, Dahl M. Analysis of DNA flanking the treA gene of Bacillus subtilis reveals genes encoding a putative specific enzyme IITre and a potential regulator of the trehalose operon. Gene [Internet]. 1996 Oct 1 [cited 2021 Sep 2];175(1–2):59–63. Available from: https://europepmc.org/article/MED/8917076

169. Kearns D, Chu F, Branda S, Kolter R, Losick R. A master regulator for biofilm formation by Bacillus subtilis. Mol Microbiol [Internet]. 2005 Feb 1 [cited 2021 Sep 2];55(3):739–49. Available from: https://europepmc.org/article/MED/15661000

170. Nogueira I, Nogueira T, Soares S, de Lencastre H. The Bacillus subtilis L-arabinose (ara) operon: nucleotide sequence, genetic organization and expression. Microbiology [Internet]. 1997 [cited 2021 Sep 2];143 (Pt 3(3):957–69. Available from: https://pubmed.ncbi.nlm.nih.gov/9084180/

171. Sirko A, Hryniewicz M, Hulanicka D, Böck A. Sulfate and thiosulfate transport in Escherichia coli K-12: nucleotide sequence and expression of the cysTWAM gene cluster. J Bacteriol [Internet]. 1990 [cited 2021 Sep 2];172(6):3351–7. Available from: https://pubmed.ncbi.nlm.nih.gov/2188958/

172. Chowdhury N, Norris J, McAlister E, Lau S, Thomas G, Boyd E. The VC1777-VC1779 proteins are members of a sialic acid-specific subfamily of TRAP transporters (SiaPQM) and constitute the sole route of sialic acid uptake in the human pathogen Vibrio cholerae. Microbiology [Internet]. 2012 May 3 [cited 2021 Sep 2];158(Pt 8):2158–67. Available from: https://europepmc.org/article/MED/22556361

173. Mathiopoulos C, Mueller J, Slack F, Murphy C, Patankar S, Bukusoglu G, et al. A Bacillus subtilis dipeptide transport system expressed early during sporulation. Mol Microbiol [Internet]. 1991 Aug 1 [cited 2021 Sep 2];5(8):1903–13. Available from: https://europepmc.org/article/MED/1766370

174. Perego M, Higgins C, Pearce S, Gallagher M, Hoch J. The oligopeptide transport system of Bacillus subtilis plays a role in the initiation of sporulation. Mol Microbiol [Internet]. 1991 [cited 2021 Sep 2];5(1):173–85. Available from: https://pubmed.ncbi.nlm.nih.gov/1901616/

175. Koide A, Hoch JA. Identification of a second oligopeptide transport system in Bacillus subtilis and determination of its role in sporulation. Mol Microbiol. 1994;13(3):417–26.

176. Parra-Lopez C, Baer MT, Groisman EA. Molecular genetic analysis of a locus required for resistance to antimicrobial peptides in Salmonella typhimurium. EMBO J [Internet]. 1993 Nov 1 [cited 2020 May 18];12(11):4053–62. Available from: http://doi.wiley.com/10.1002/j.1460-2075.1993.tb06089.x

177. Breazeale SD, Ribeiro AA, Raetz CRH. Origin of Lipid A Species Modified with 4-Amino-4-deoxy-l-arabinose in Polymyxin-resistant Mutants of Escherichia coli: AN AMINOTRANSFERASE (ArnB) THAT GENERATES UDP-4-AMINO-4-DEOXY-l-ARABINOSE. J Biol Chem. 2003 Jul 4;278(27):24731–9.

178. Yan A, Guan Z, Raetz CRH. An Undecaprenyl Phosphate-Aminoarabinose Flippase Required for Polymyxin Resistance in Escherichia coli. J Biol Chem. 2007 Dec 7;282(49):36077–89.

179. Liu YY, Wang Y, Walsh TR, Yi LX, Zhang R, Spencer J, et al. Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: A microbiological and molecular biological study. Lancet Infect Dis. 2016

Feb 1;16(2):161–8.

180. Alonso A, Sanchez P, Martínez JL. Stenotrophomonas maltophilia D457R Contains a Cluster of Genes from Gram-Positive Bacteria Involved in Antibiotic and Heavy Metal Resistance. Antimicrob Agents Chemother [Internet]. 2000 Jul [cited 2021 Nov 9];44(7):1778. Available from: /pmc/articles/PMC89961/

181. Humphrey S, Fillol-Salom A, Quiles-Puchalt N, Ibarra-Chávez R, Haag AF, Chen J, et al. Bacterial chromosomal mobility via lateral transduction exceeds that of classical mobile genetic elements. Nat Commun 2021 121 [Internet]. 2021 Nov 8 [cited 2021 Nov 9];12(1):1–13. Available from: https://www.nature.com/articles/s41467-021-26004-5

182. Chiang YN, Penadés JR, Chen J. Genetic transduction by phages and chromosomal islands: The new and noncanonical. PLoS Pathog [Internet]. 2019 [cited 2021 Nov 9];15(8). Available from: /pmc/articles/PMC6687093/

183. Royal College of Physicians I, Health Service Executive I. Guidelines for the Prevention and Control of MDRO excluding MRSA in the healthcare setting. 2012;

184. Wein T, Wang Y, Barz M, Stücker FT, Hammerschmidt K, Dagan T. Essential gene acquisition destabilizes plasmid inheritance. PLOS Genet [Internet]. 2021 Jul 12 [cited 2021 Aug 19];17(7):e1009656. Available from: https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1009656

185. Alice AF, López CS, Crosa JH. Plasmid- and Chromosome-Encoded Redundant and Specific Functions Are Involved in Biosynthesis of the Siderophore Anguibactin in Vibrio anguillarum 775: a Case of Chance and Necessity? J Bacteriol [Internet]. 2005 Mar [cited 2021 Aug 19];187(6):2209. Available from: /pmc/articles/PMC1064064/

186. Zheng J, Guan Z, Cao S, Peng D, Ruan L, Jiang D, et al. Plasmids are vectors for redundant chromosomal genes in the Bacillus cereus group. BMC Genomics 2015 161

[Internet]. 2015 Jan 22 [cited 2021 Aug 19];16(1):1–10. Available from: https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-014-1206-5

187. Baker-Austin C, Wright MS, Stepanauskas R, McArthur J V. Co-selection of antibiotic and metal resistance. Trends Microbiol. 2006 Apr;14(4):176–82.

188. Pal C, Asiani K, Arya S, Rensing C, Stekel D, Larsson D, et al. Metal Resistance and Its Association With Antibiotic Resistance. Adv Microb Physiol [Internet]. 2017 Jan 1 [cited 2021 Aug 18];70:261–313. Available from: https://pubmed.ncbi.nlm.nih.gov/28528649/

189. Bae T, Baba T, Hiramatsu K, Schneewind O. Prophages of Staphylococcus aureus Newman and their contribution to virulence. Mol Microbiol. 2006 Nov;62(4):1035–47.

190. Xia G, Wolz C. Phages of Staphylococcus aureus and their impact on host evolution. Infect Genet Evol. 2014 Jan 1;21:593–601.

191. Pasek S, Risler J-L, Brezellec P. Gene fusion/fission is a major contributor to evolution of multi-domain bacterial proteins. Bioinformatics. 2006 Jun;22(12):1418–23.

192. Des Marais DL, Rausher MD. Escape from adaptive conflict after duplication in an anthocyanin pathway gene. Nature [Internet]. 2008 Aug 7 [cited 2019 Feb 11];454(7205):762–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18594508

193. Rastogi S, Liberles DA. Subfunctionalization of duplicated genes as a transition state to neofunctionalization. BMC Evol Biol [Internet]. 2005 Apr [cited 2019 Feb 19];5(1):28. Available from: http://bmcevolbiol.biomedcentral.com/articles/10.1186/1471-2148-5-28

194. Leonard G, Richards TA. Genome-scale comparative analysis of gene fusions, gene fissions, and the fungal tree of life. Proc Natl Acad Sci [Internet]. 2012 Dec 26 [cited

2017 Mar 7];109(52):21402–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23236161

195. Conant GC, Birchler JA, Pires JC. Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time. Curr Opin Plant Biol. 2014 Jun;19:91–8.

196. Wang W, Yu H, Long M. Duplication-degeneration as a mechanism of gene fission and the origin of new genes in Drosophila species. Nat Genet [Internet]. 2004 May 4 [cited 2017 Jul 18];36(5):523–7. Available from: http://www.nature.com/doifinder/10.1038/ng1338

197. Snel B, Bork P, Huynen M. Genome evolution. Gene fusion versus gene fission. Trends Genet [Internet]. 2000 Jan [cited 2017 Jul 18];16(1):9–11. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10637623

198. De Been M, Van Schaik W, Cheng L, Corander J, Willems RJ. Recent recombination events in the core genome are associated with adaptive evolution in Enterococcus faecium. Genome Biol Evol [Internet]. 2013 Aug 1 [cited 2021 Nov 12];5(8):1524–35. Available from: https://pubmed.ncbi.nlm.nih.gov/23882129/

199. Leavis HL, Willems RJL, Van Wamel WJB, Schuren FH, Caspers MPM, Bonten MJM. Insertion sequence-driven diversification creates a globally dispersed emerging multiresistant subspecies of E. faecium. PLoS Pathog. 2007 Jan;3(1):0075–96.

200. Arenas M, Araujo NM, Branco C, Castelhano N, Castro-Nallar E, Pérez-Losada M. Mutation and recombination in pathogen evolution: Relevance, methods and controversies. Infect Genet Evol [Internet]. 2018 Sep 1 [cited 2021 Nov 12];63:295–306. Available from: https://pubmed.ncbi.nlm.nih.gov/28951202/

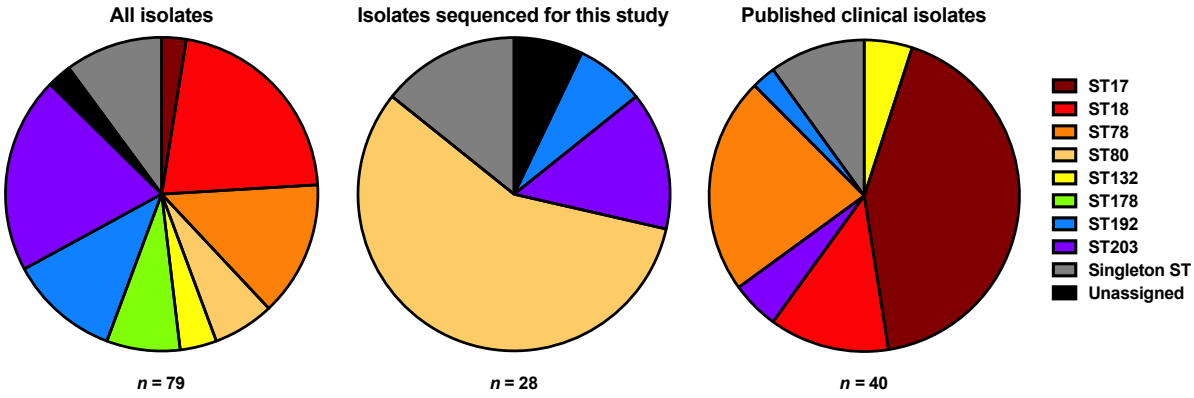


Figure 1: Distribution of sequence types between different studies

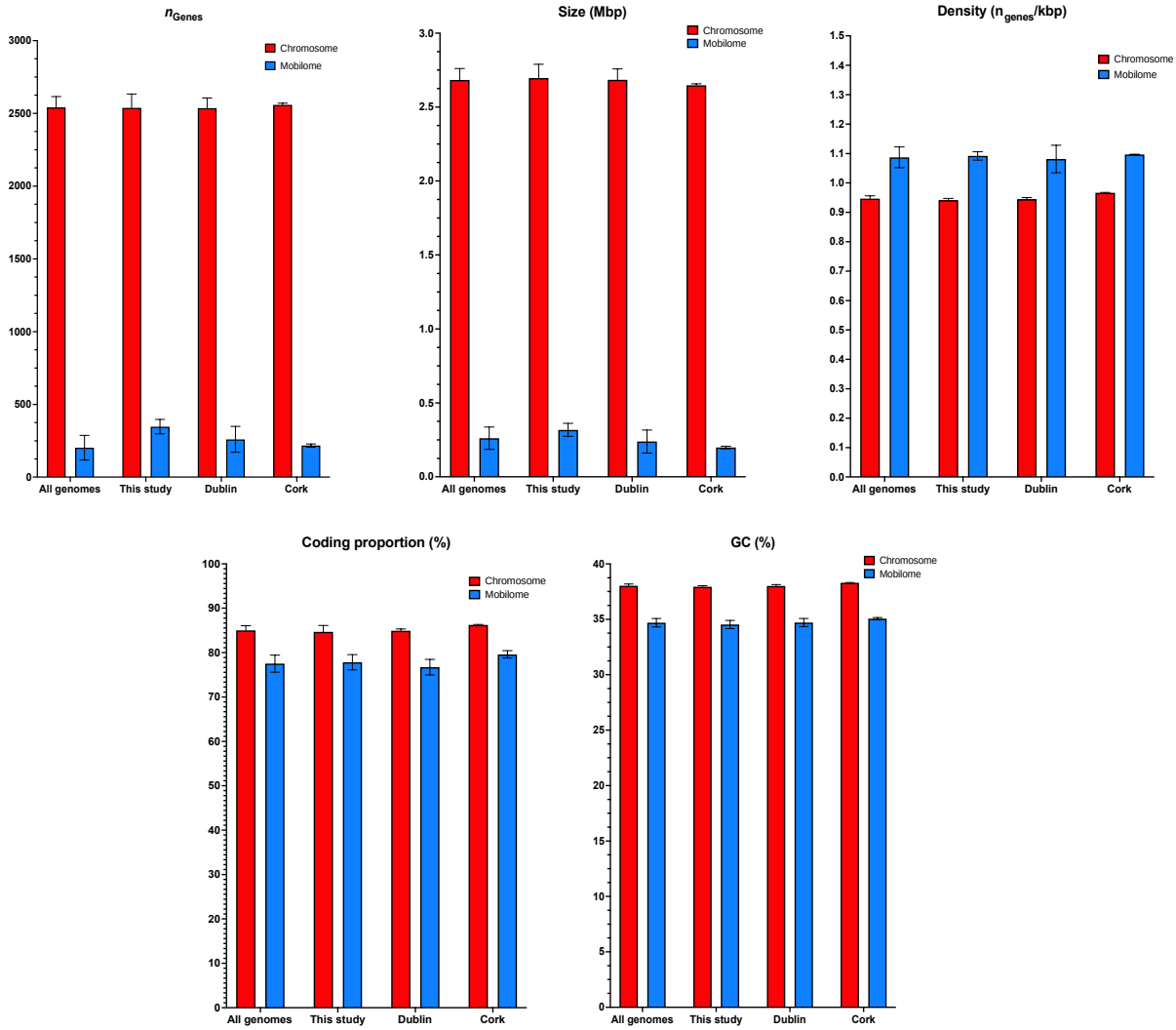


Figure 2: Comparison of different genome characteristics between chromosomes and plasmids across different studies.

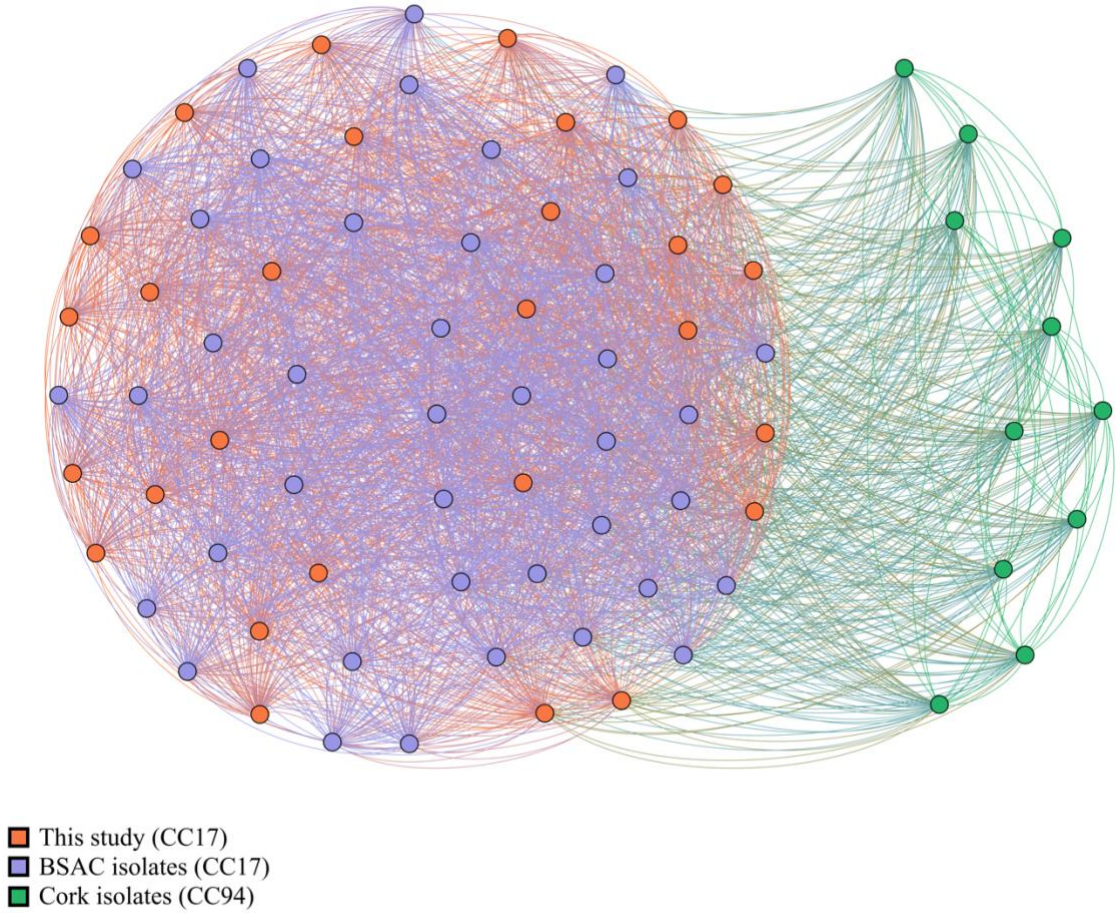


Figure 3: Genomic distance cluster analysis of chromosomal sequences from all isolates. All CC17 isolates cluster together and all CC94 isolates cluster together (whereby all isolates share an edge). Approximately half of CC17 cluster with CC94.

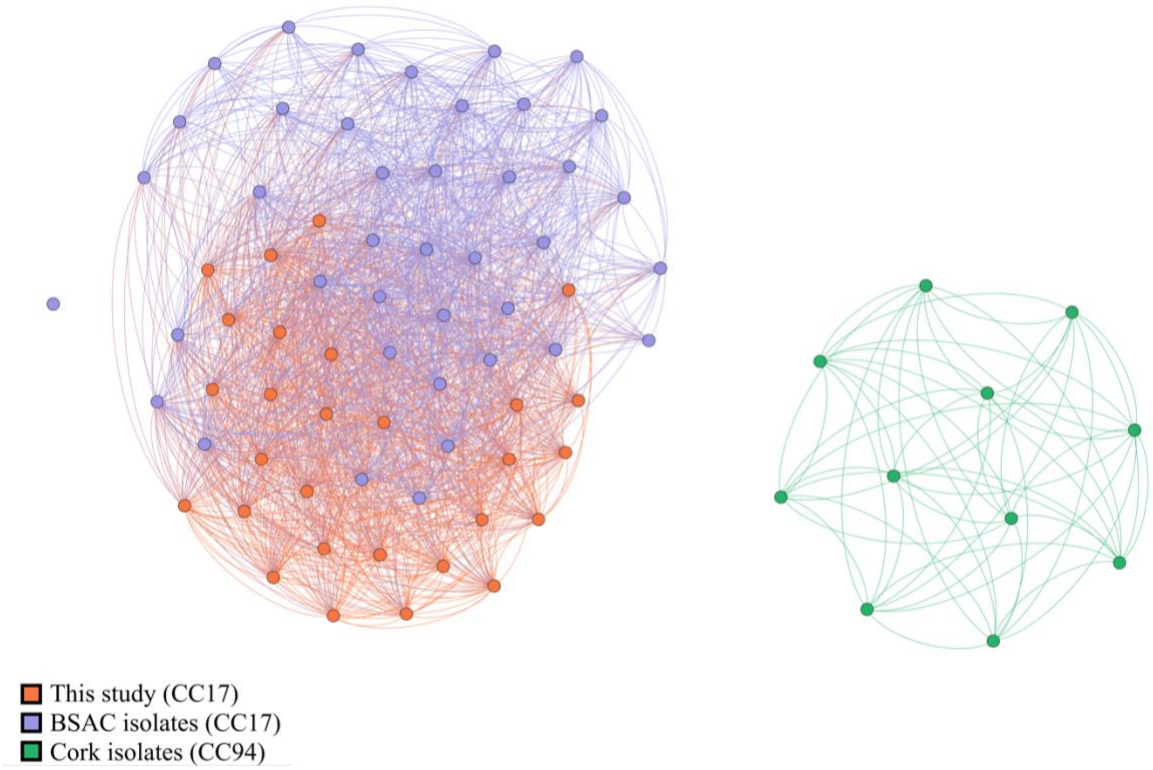


Figure 4: Genomic distance cluster analysis of mobilomal sequences from all isolates. While some mixing occurs with BSAC isolates and isolates sequenced for this study, divergence between communities is observed. One BSAC isolate (ERR374724) is completely disconnected from the other connected components, which is unusual due to the shared presence of a *vanA*+ genotype between ERR374724 and many other isolates. All CC94 isolates cluster together in a separate connected component from CC17 mobilomes.

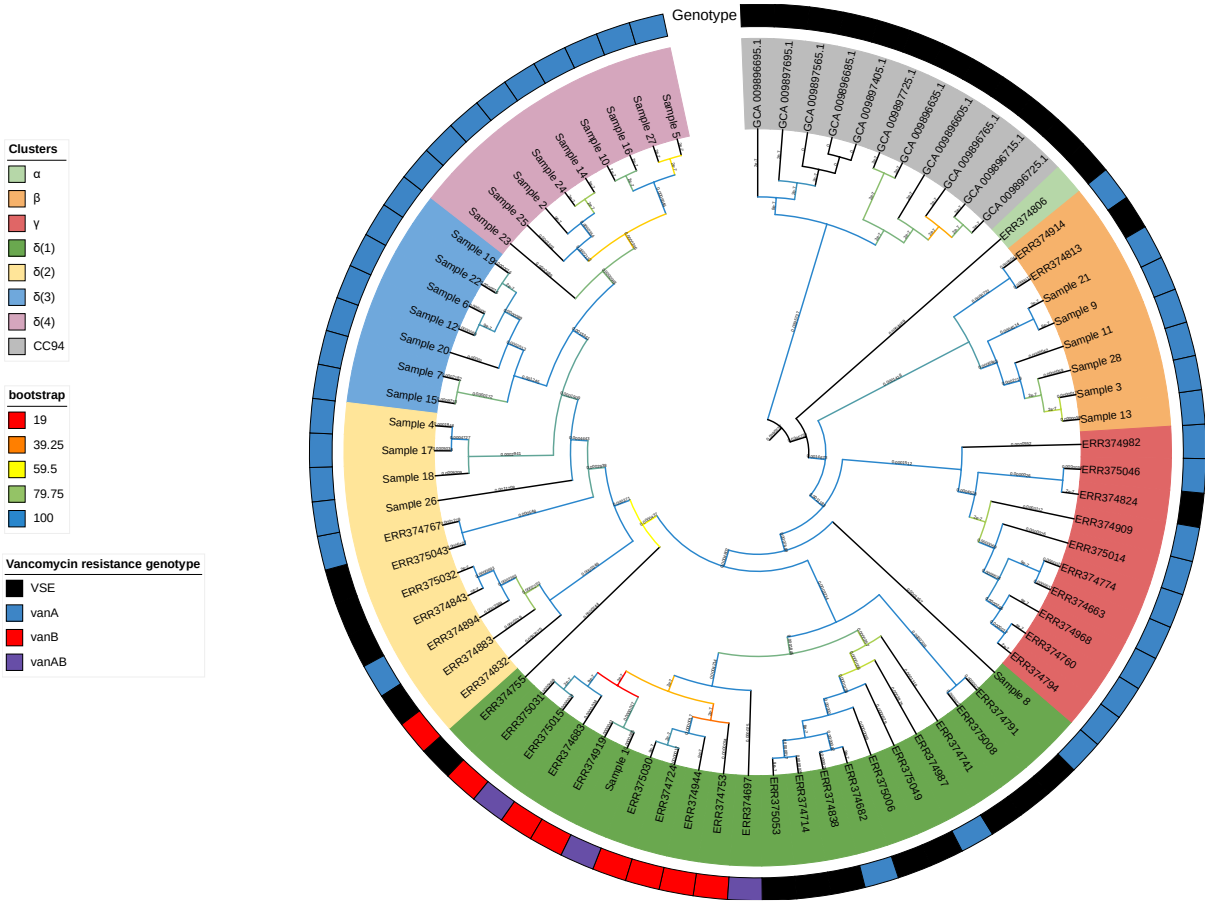


Figure 5: A 10,000 bootstrap phylogeny of isolates. Taxa are shaded based on their RheirBAPS clustering (discussed in methodology section) with an outer ring denoting the associated vancomycin resistance genotypes for each taxon. Bootstrap supports are provided for internal nodes. The majority of supports equate to 100%. Bootstrap shading is transitional, the colours shown in the legend are equidistant scales denoting where a new colour is shown. For example, a datapoint between 79.25% and 100% would be a shade of turquoise. Deep nodes all received high bootstrap support, further corroborating RheirBAPS clustering.

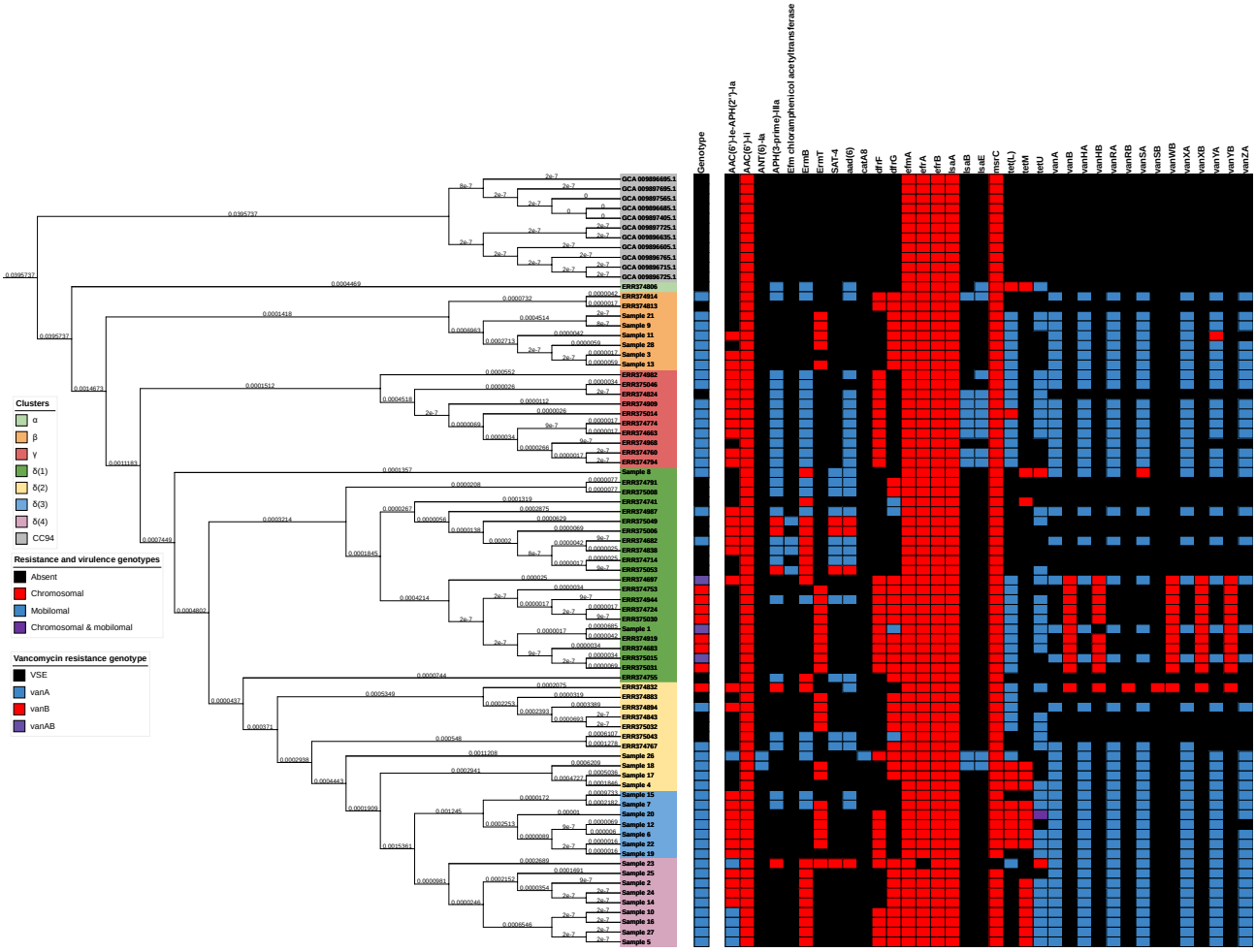


Figure 6. A phylogeny of isolates illustrating drug resistance genotype distributions as a categorical heatmap. Again, taxa are shaded based on their associated RheirBAPS cluster and a vancomycin resistance genotype is given as a separate bar for each taxon.

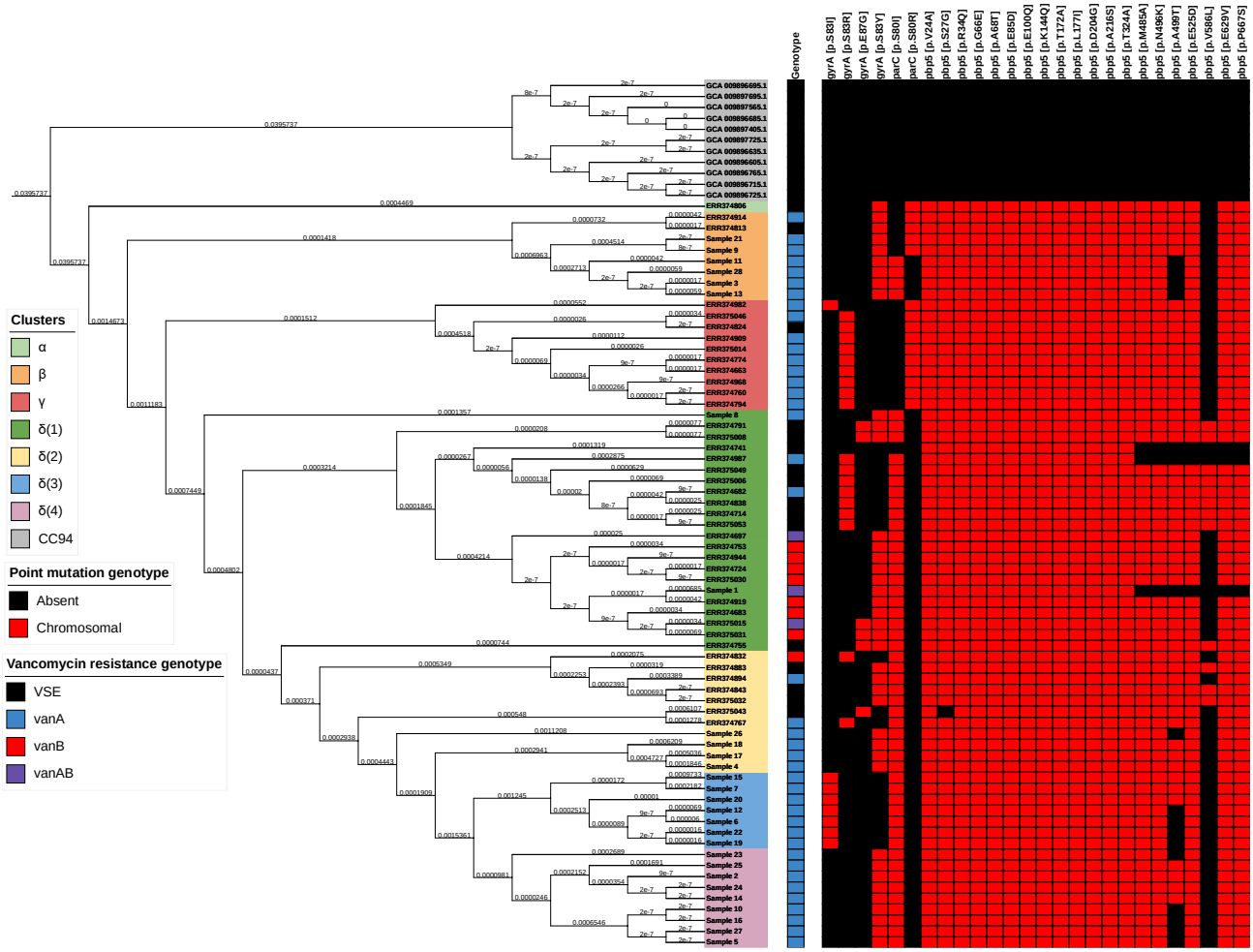


Figure 7. A phylogeny of isolates illustrating point mutation mediated drug resistance genotype distributions as a categorical heatmap. Again, taxa are shaded based on their associated RheirBAPS cluster and a vancomycin resistance genotype is given as a separate bar for each taxon.

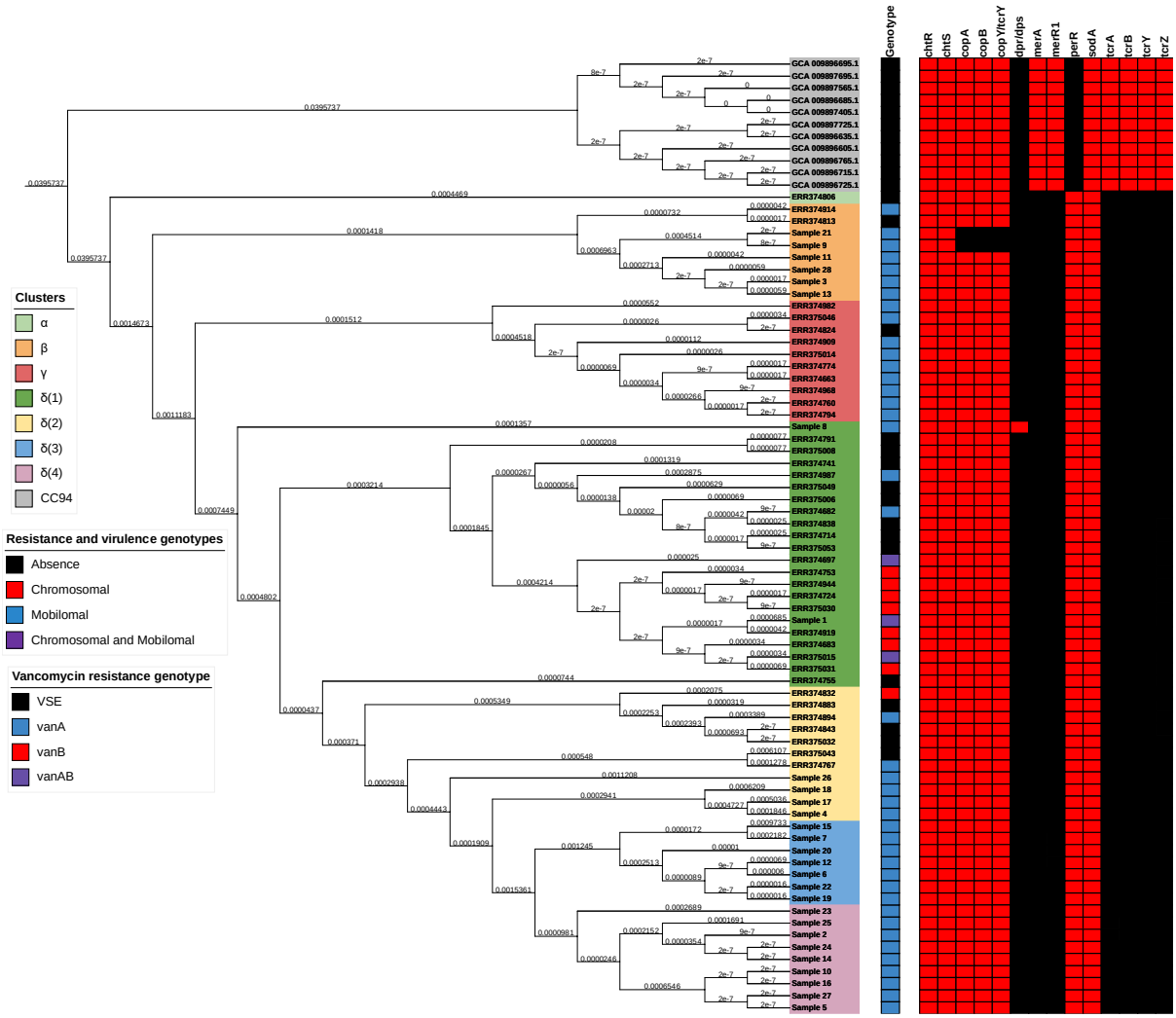


Figure 8. A phylogeny of isolates illustrating metal (biocide) resistance genotype distributions as a categorical heatmap. Again, taxa are shaded based on their associated RheirBAPS cluster and a vancomycin resistance genotype is given as a separate bar for each taxon.

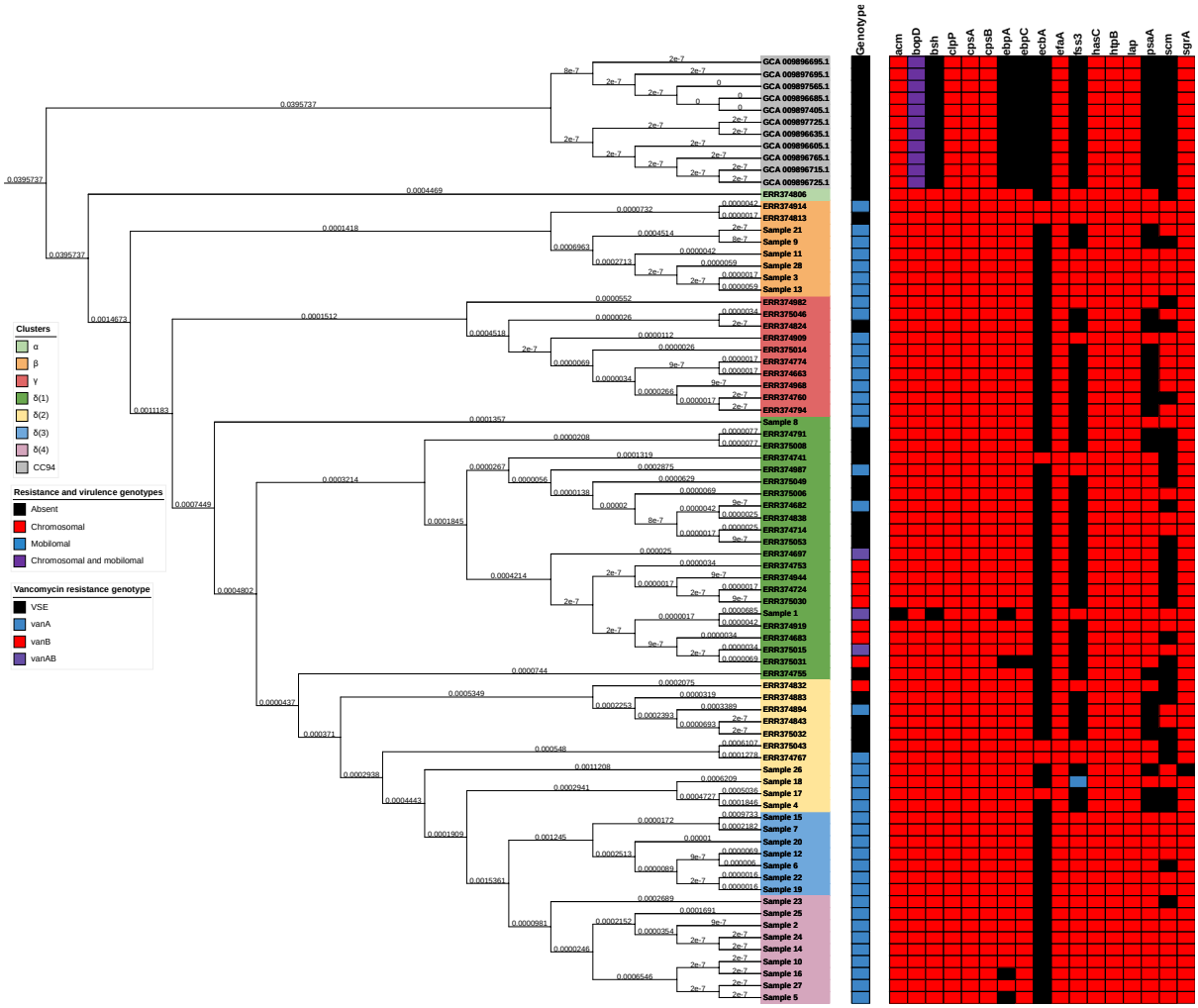


Figure 9. A phylogeny of isolates illustrating virulence factor genotype distributions as a categorical heatmap. Again, taxa are shaded based on their associated RheirBAPS cluster and a vancomycin resistance genotype is given as a separate bar for each taxon.

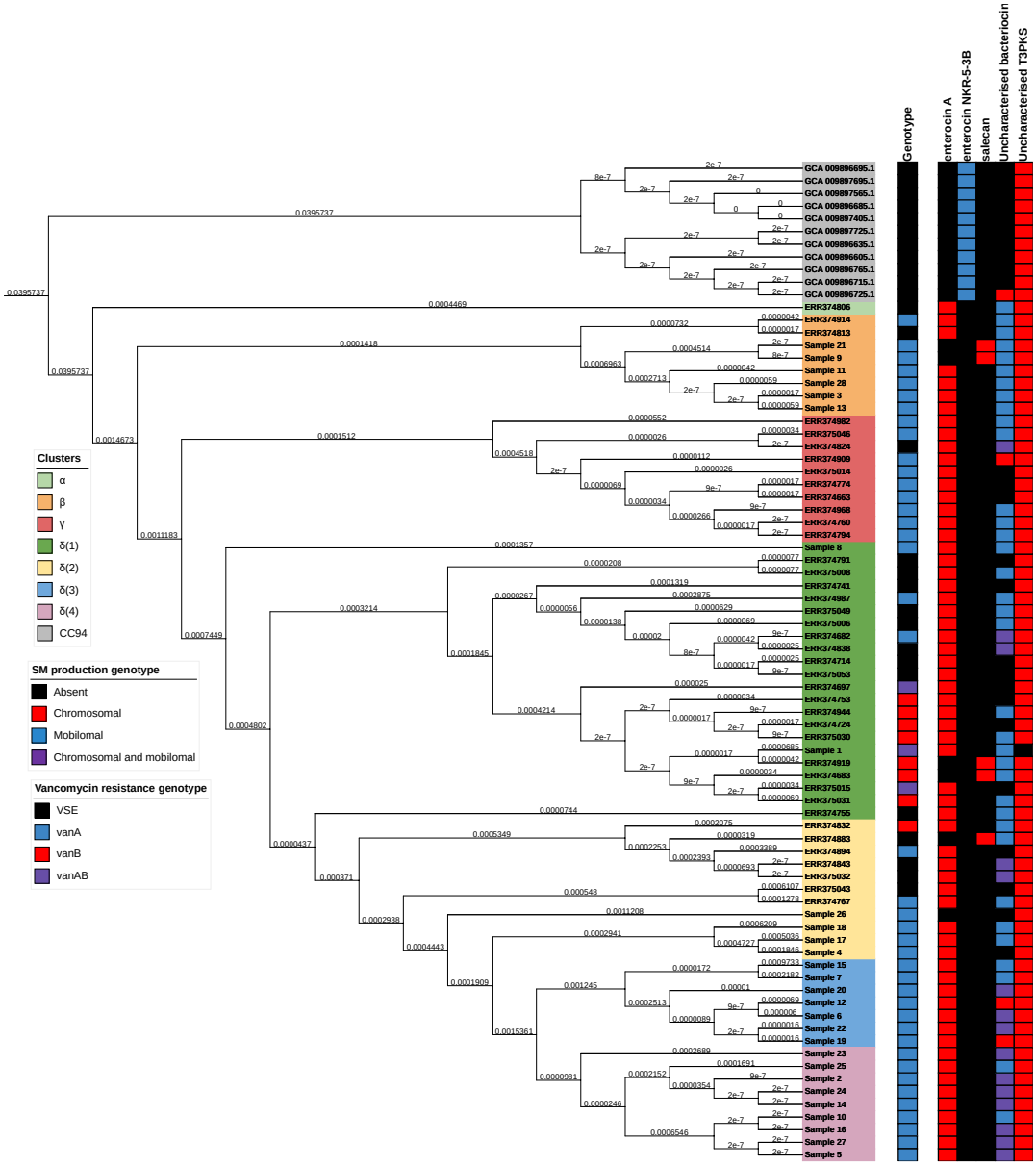


Figure 10. A phylogeny of isolates illustrating secondary metabolite production genotype distributions as a categorical heatmap. Again, taxa are shaded based on their associated RheirBAPS cluster and a vancomycin resistance genotype is given as a separate bar for each taxon.

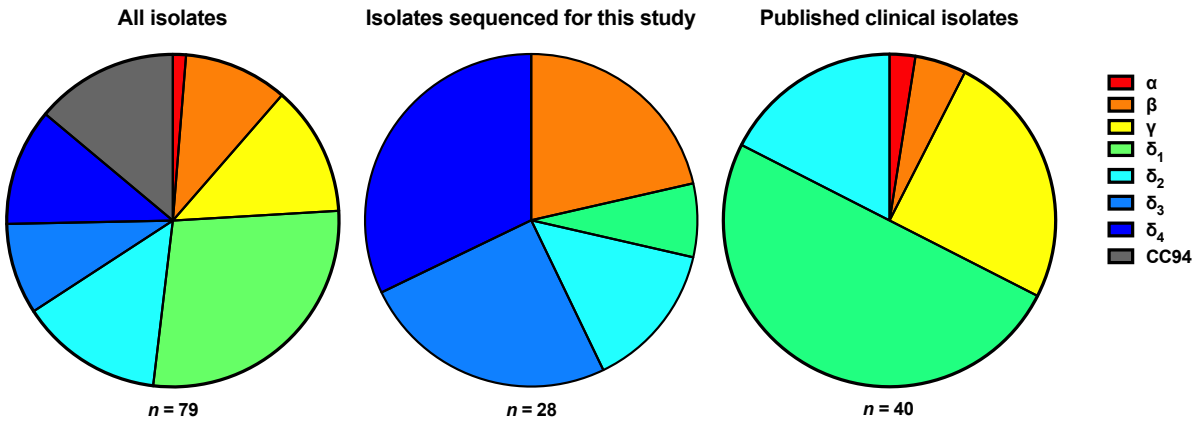


Figure 11: Distribution of RheirBAPS clusters between studies

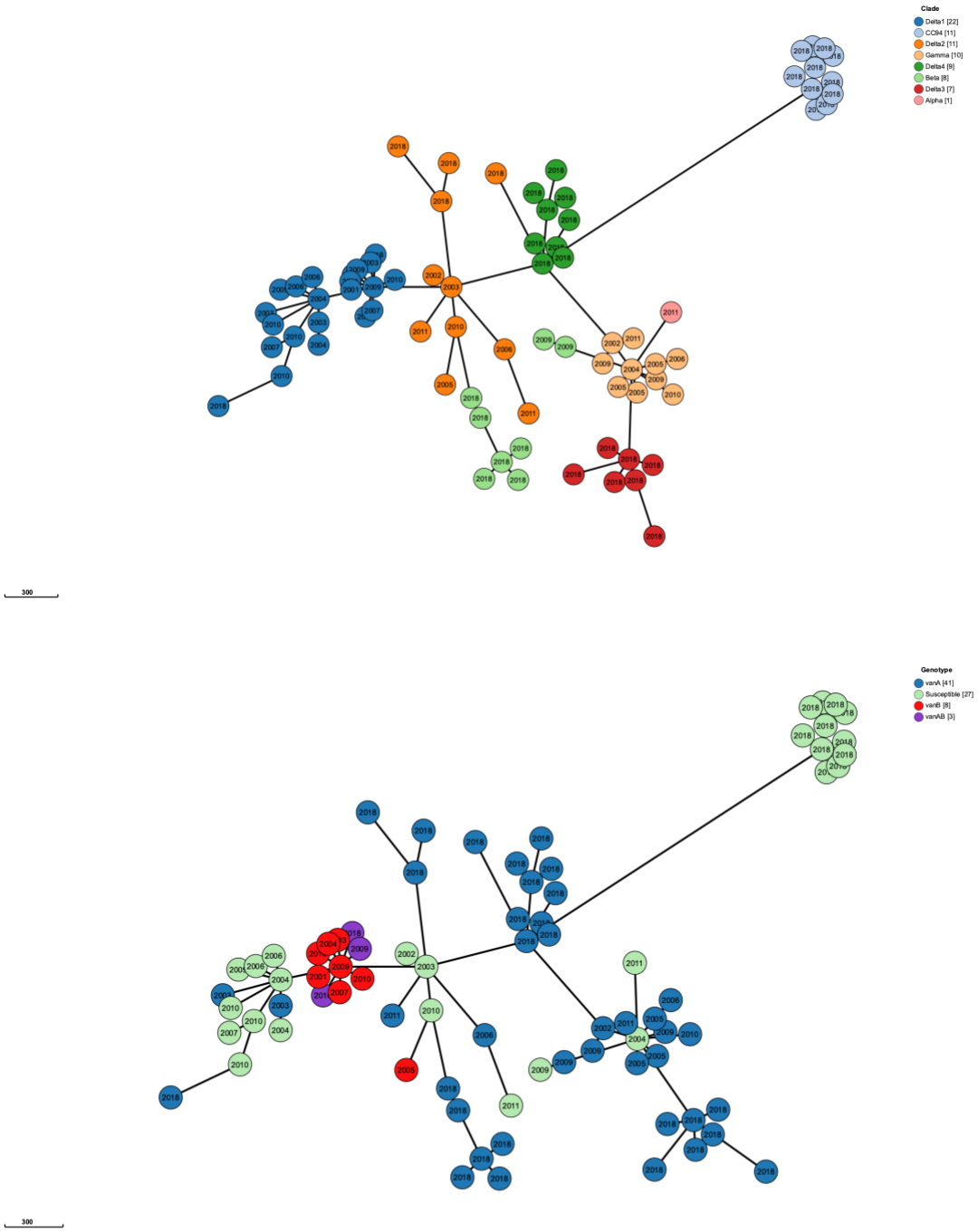


Figure 12: Core genome MLST (cgMLST) for all isolates included in this study. The upper image is shaded based on RheirBAPS clustering and the lower is shaded based on vancomycin resistance genotypes. The year of isolation is presented in the centre of each circle, where each circle represents an isolate.

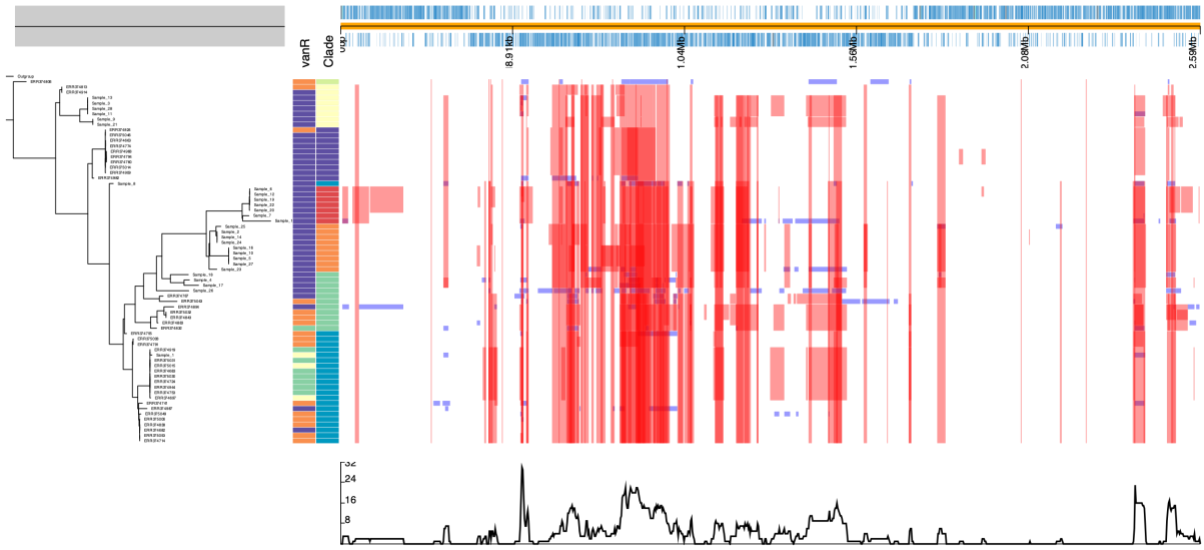


Figure 13: Red bars illustrate hot spots compared to a reference, where a hotspot is defined as a site where more than one species reports a ≥50% likelihood for the same recombination event. Blue boxes illustrate a recombination event for a single species. Multiple different recombination events can occur at the same site (appearing as overlap in this image). The plot at the bottom shows the maximal amount of species affected by a detected recombination event in a given hotspot

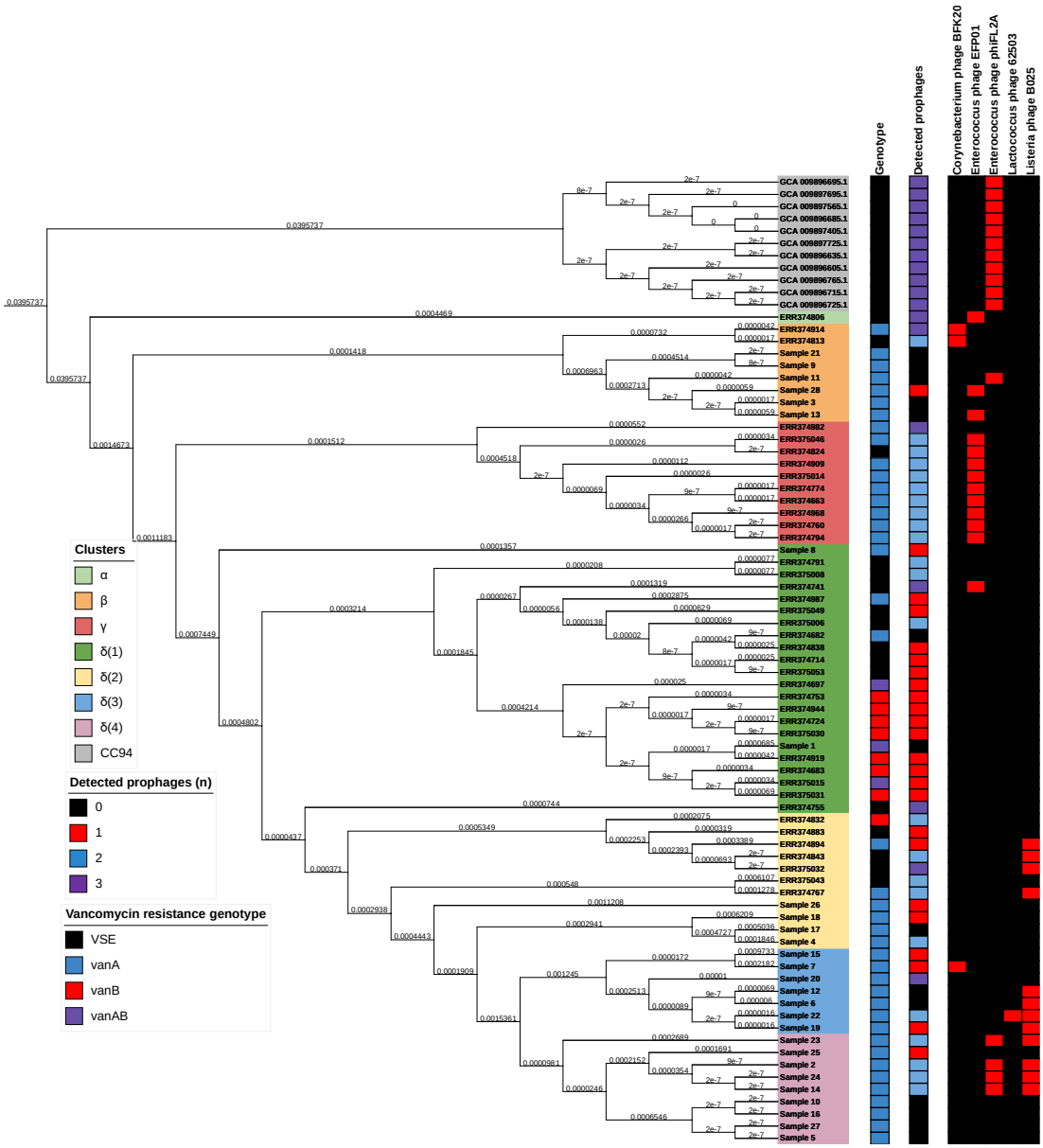


Figure 14. A phylogeny of isolates illustrating detected prophage distributions as a categorical heatmap. Again, taxa are shaded based on their associated RheirBAPS cluster and a vancomycin resistance genotype is given as a separate bar for each taxon.

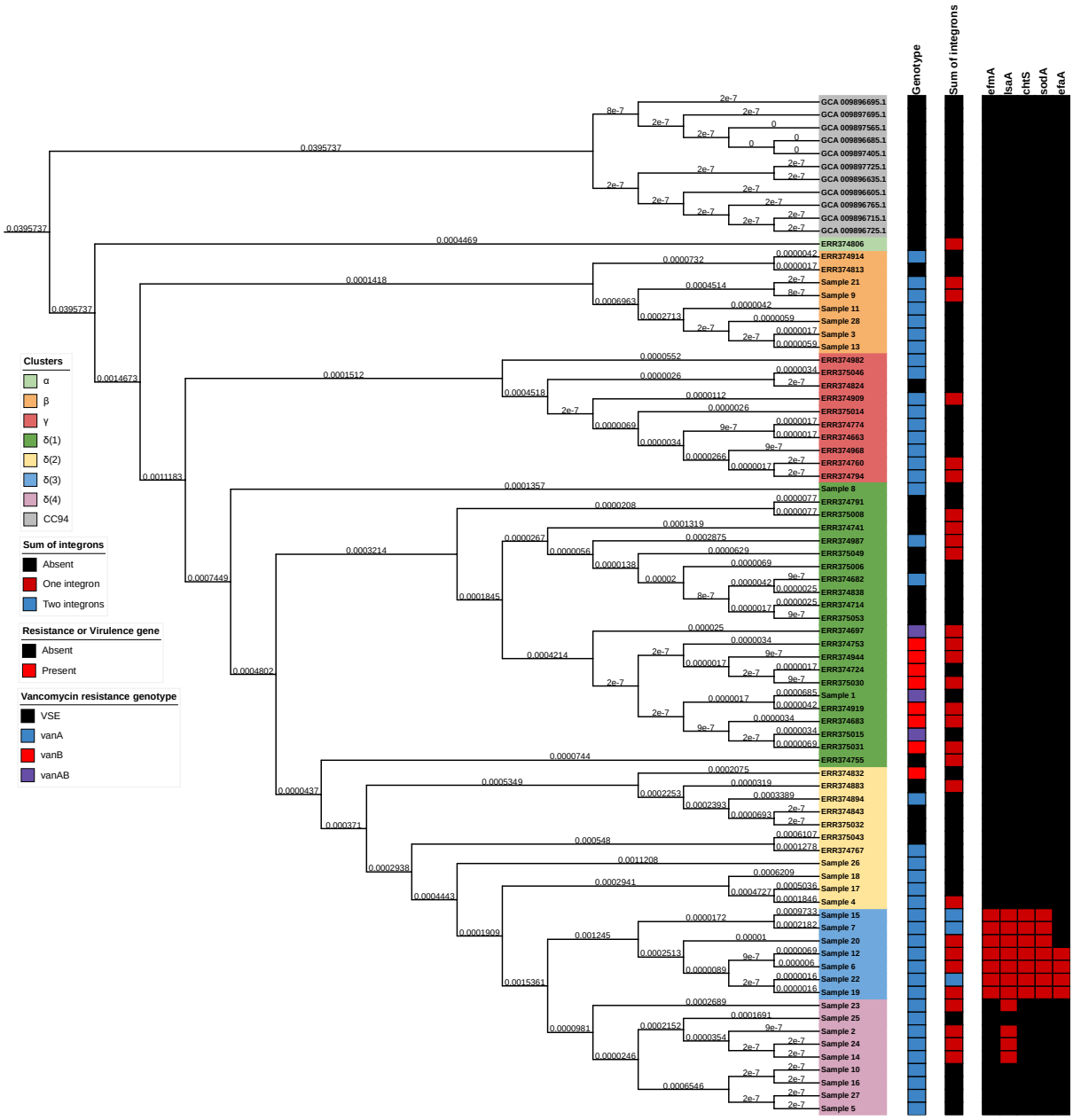
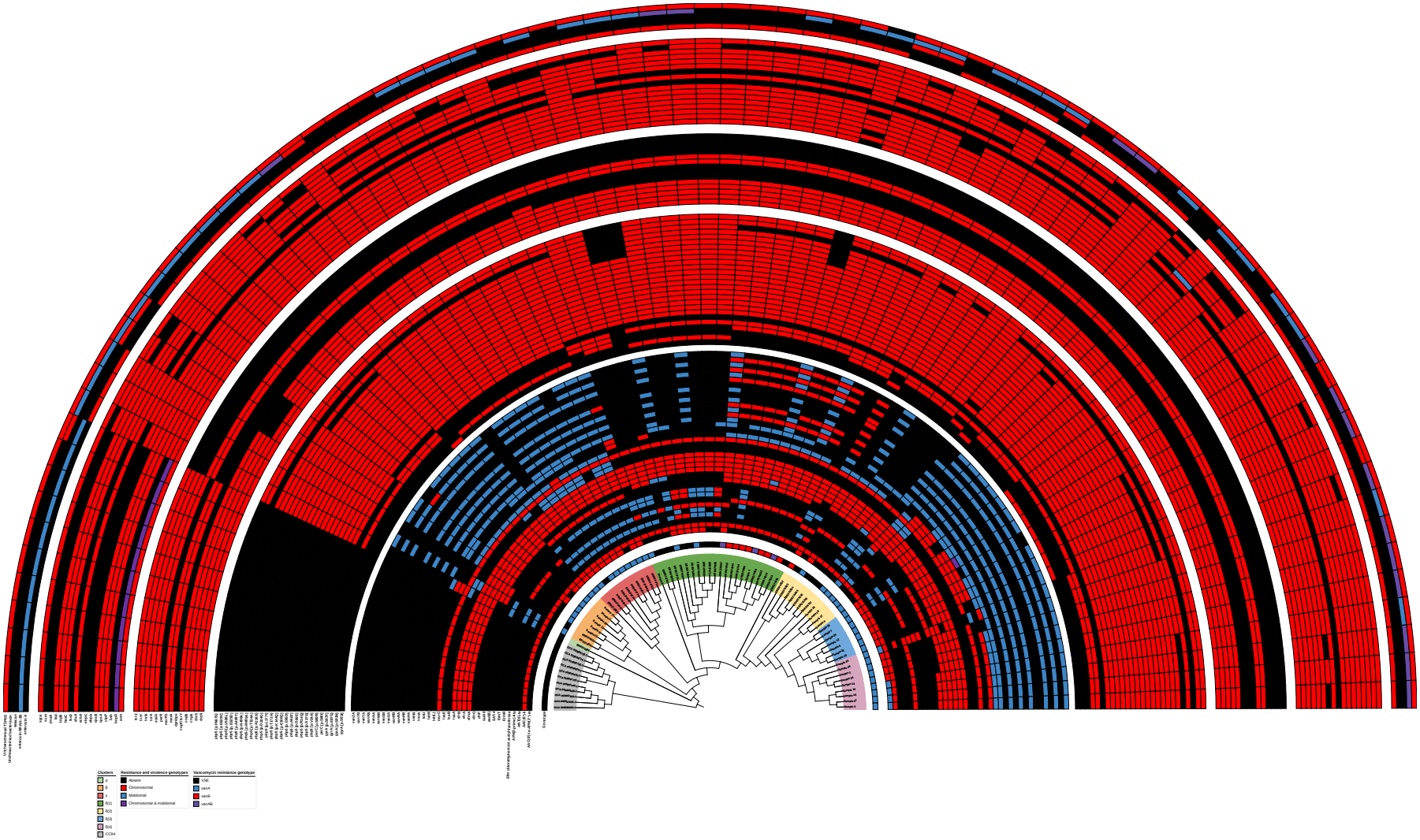
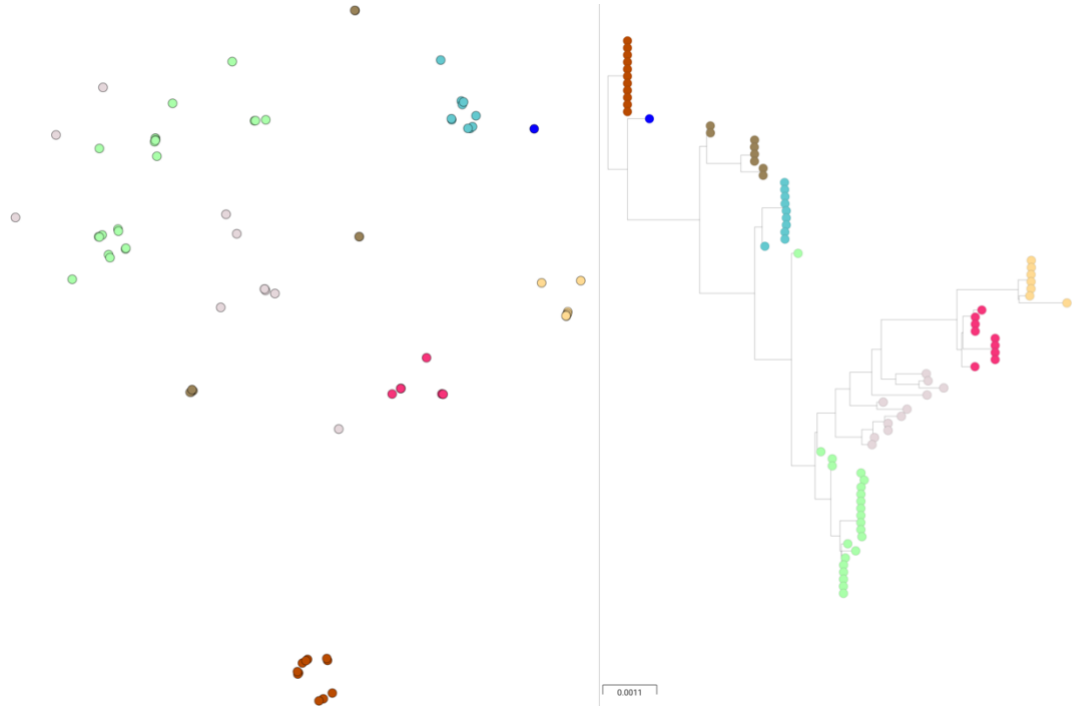


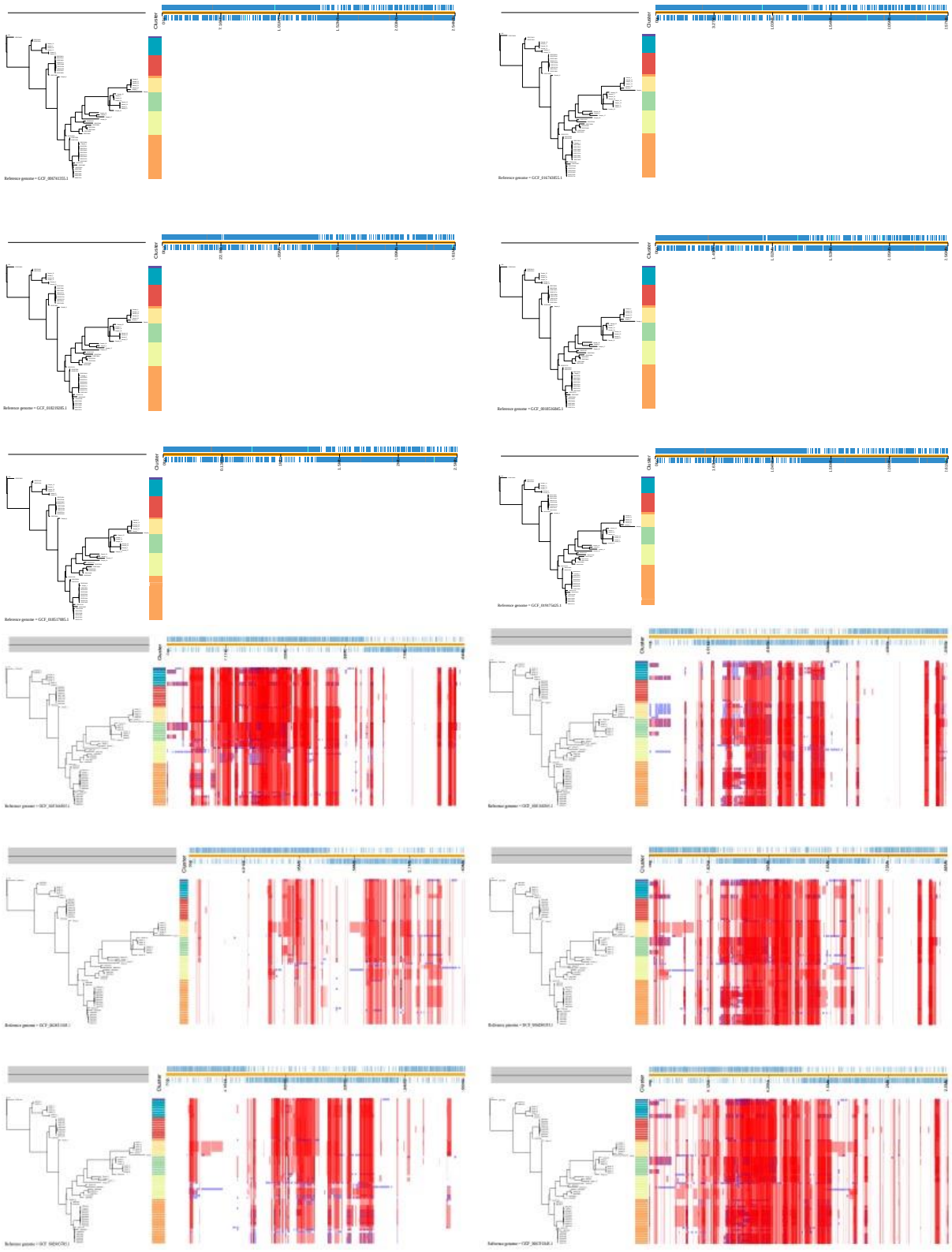
Figure 15. A phylogeny of isolates illustrating detected integron distributions as a categorical heatmap. Again, taxa are shaded based on their associated RheirBAPS cluster and a vancomycin resistance genotype is given as a separate bar for each taxon. Gene distributions (presence or absence) are given for *efmA* and *lsaA* (drug resistance), *chtS* and *sodA* (metal and biocide resistance), and *efaA* (virulence factor).

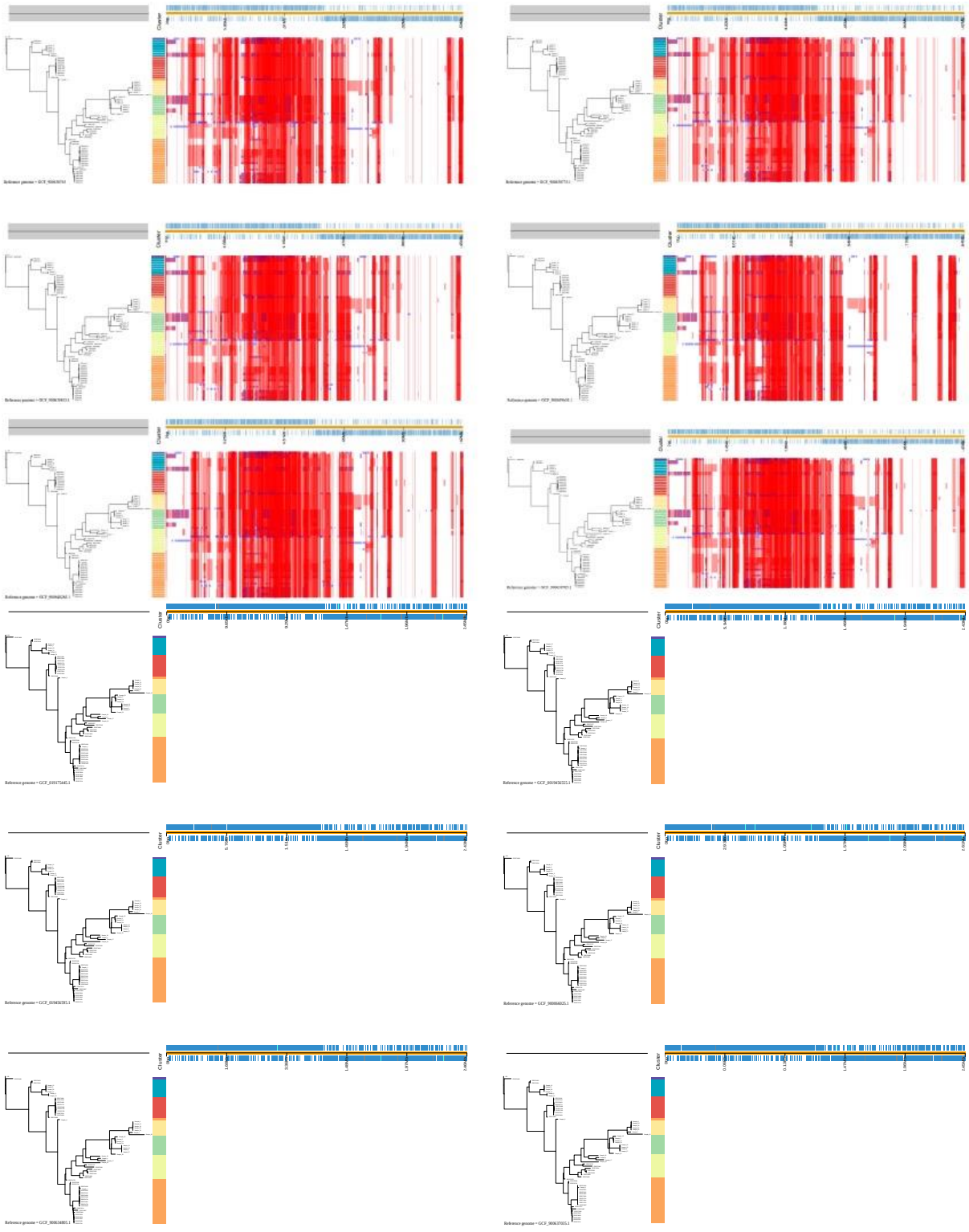


SI Figure 1: A 10,000 bootstrap phylogeny with all clinically relevant genes arranged as semi-circular categorical heatmaps. Annotation semi-circles are arranged as follows (innermost to outermost): drug resistance genotype, point mutation drug resistance genotype, metal (biocide) resistance genotype, virulence factor genotype, secondary metabolism production genotype



SI Figure 2: Core genome clustering. These clusters were produced using PANINI to corroborate cgMLST data, the phylogeny on the right is the same phylogeny used in all other images, colours are RheirBAPS clusters





SI Figure 3: Genome recombination between 24 different reference genomes. Each reference genome diverged prior to CC17. Red bars illustrate hot spots compared to a reference, where a hotspot is defined as a site where more than one species reports a ≥50% likelihood for the same recombination event. Blue boxes illustrate a recombination event for a single species. Multiple different recombination events can occur at the same site (appearing as overlap in this image). The plot at the bottom shows the maximal amount of species affected by a detected recombination event in a given hotspot. Each reference genome is given in the bottom left.

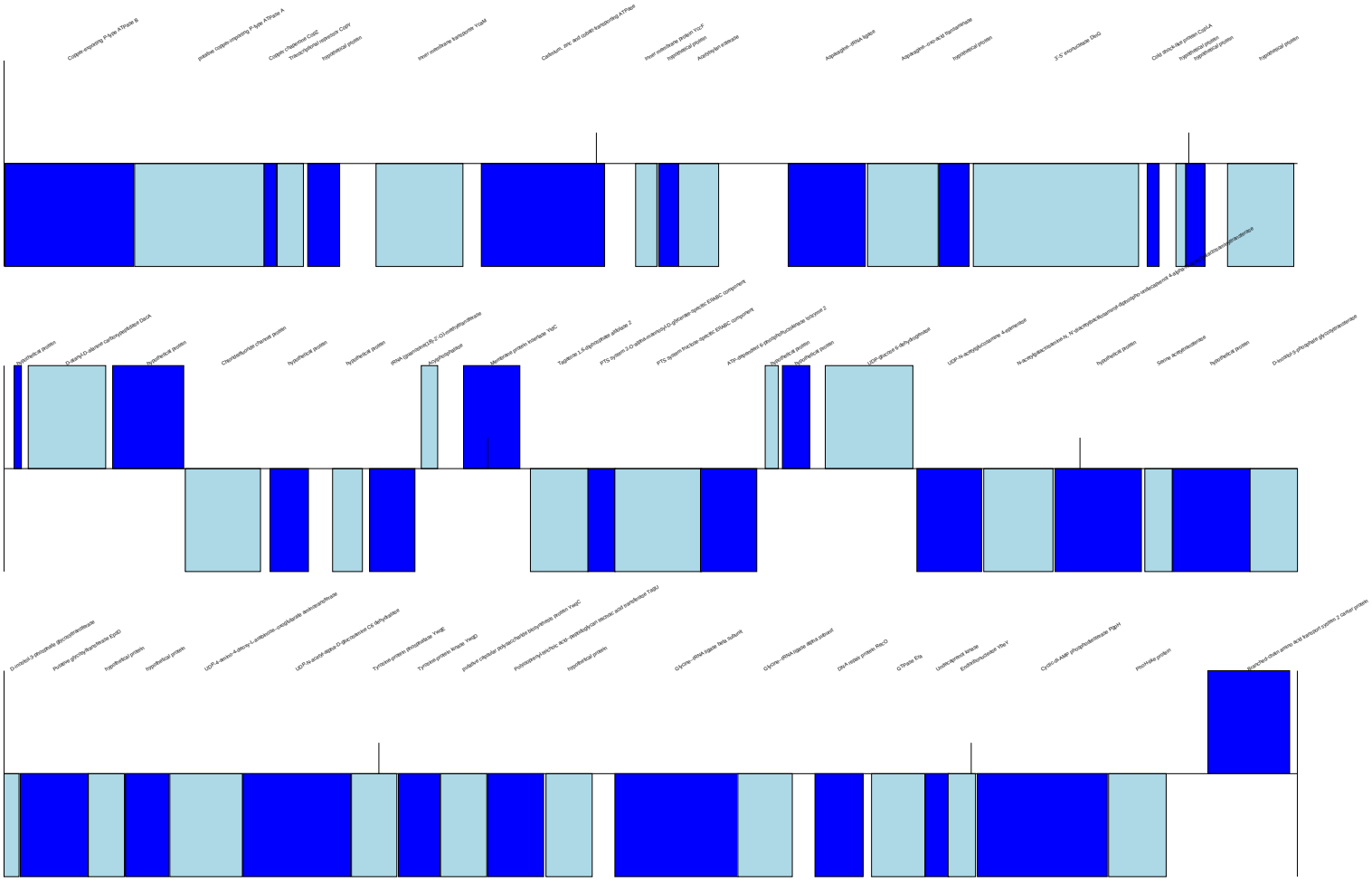


Figure 4(*a*): *arnB* contig in Sample 2. Positive sense genes are given above the line and negative sense are below the line. Genes are coloured in alternating blue for ease of visualisation

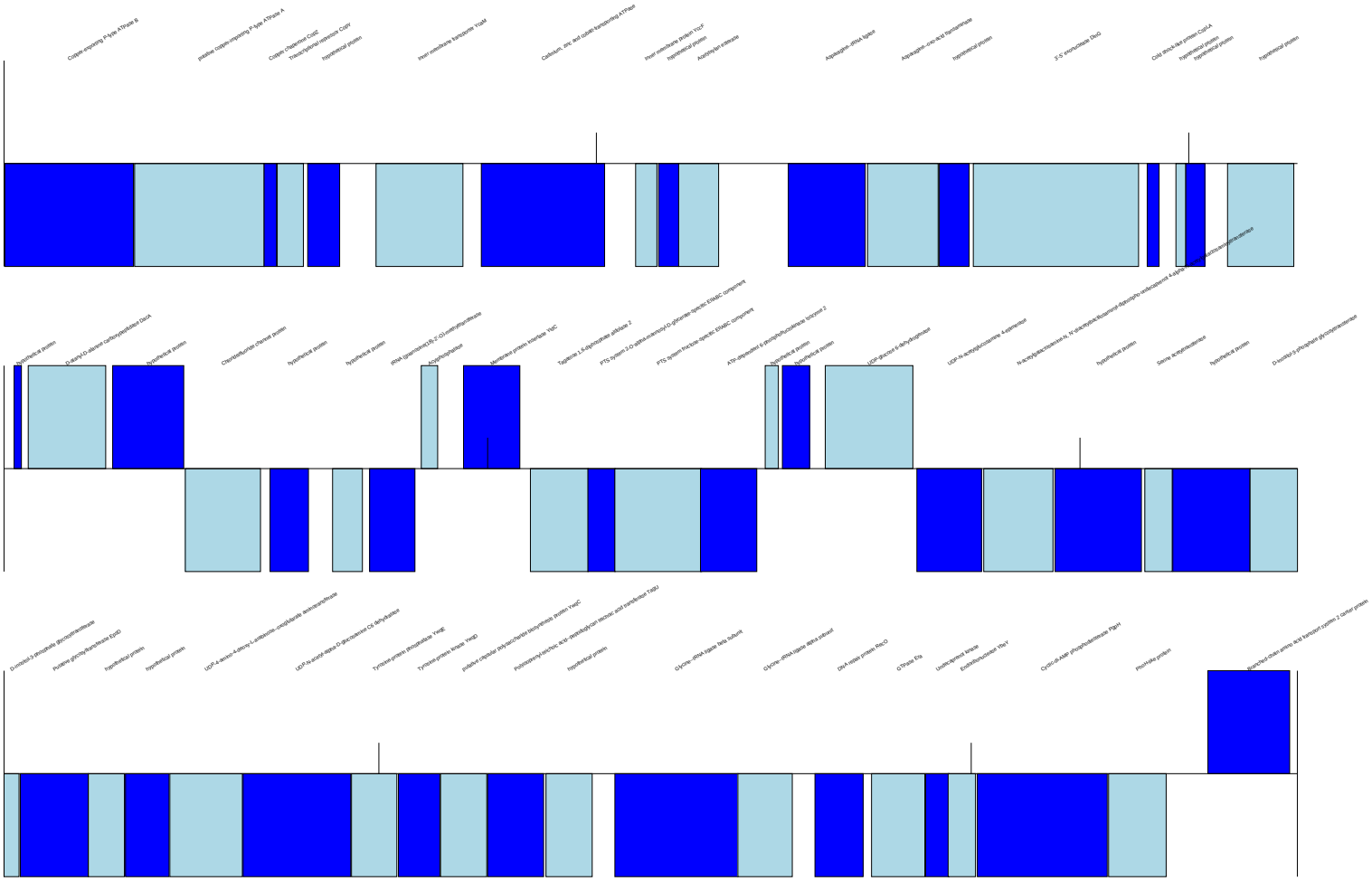


Figure 4(*b*): *arnB* contig in Sample 14. Positive sense genes are given above the line and negative sense are below the line. Genes are coloured in alternating blue for ease of visualisation

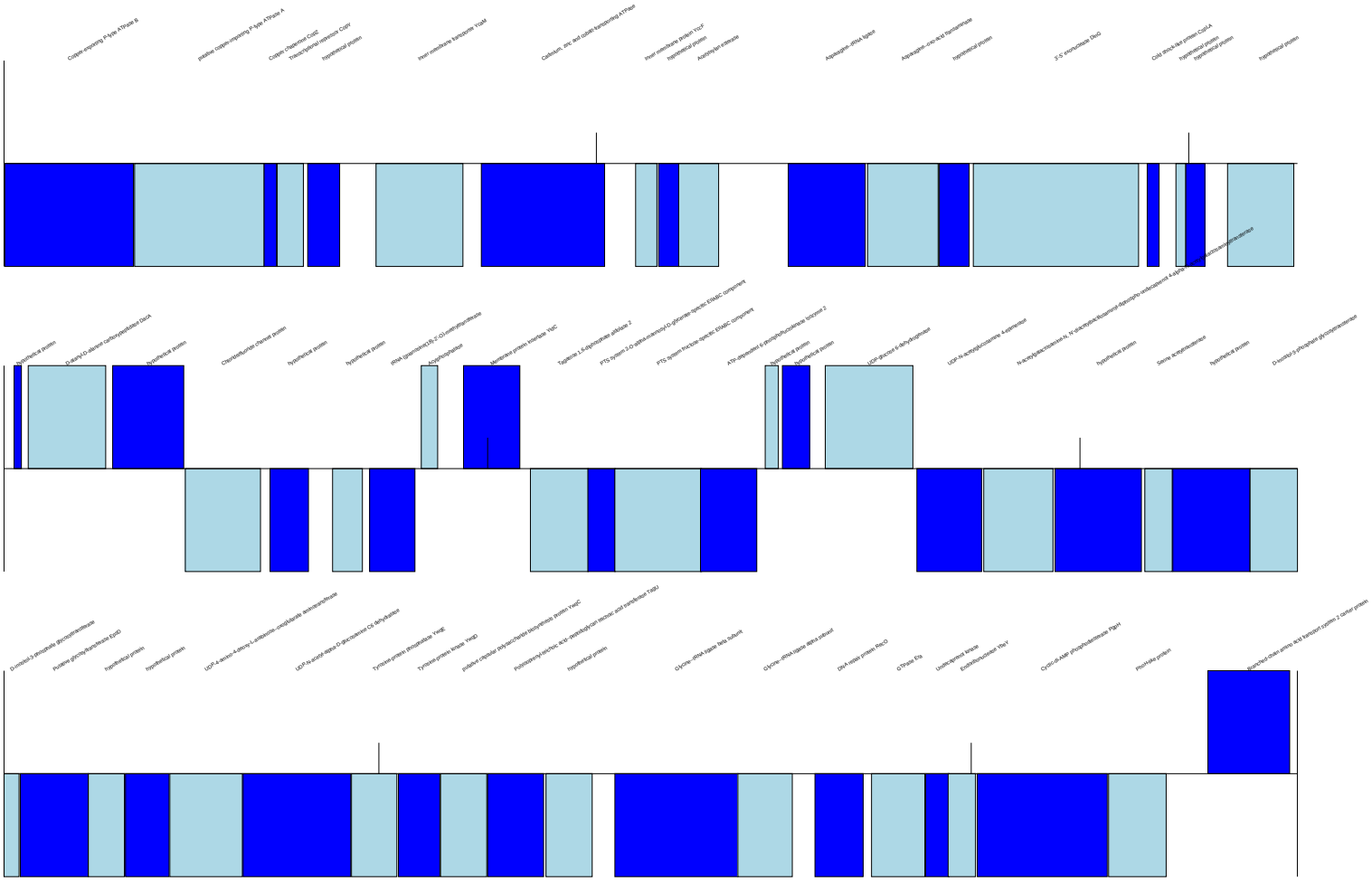


Figure 4(*c*): *arnB* contig in Sample 24. Positive sense genes are given above the line and negative sense are below the line. Genes are coloured in alternating blue for ease of visualisation

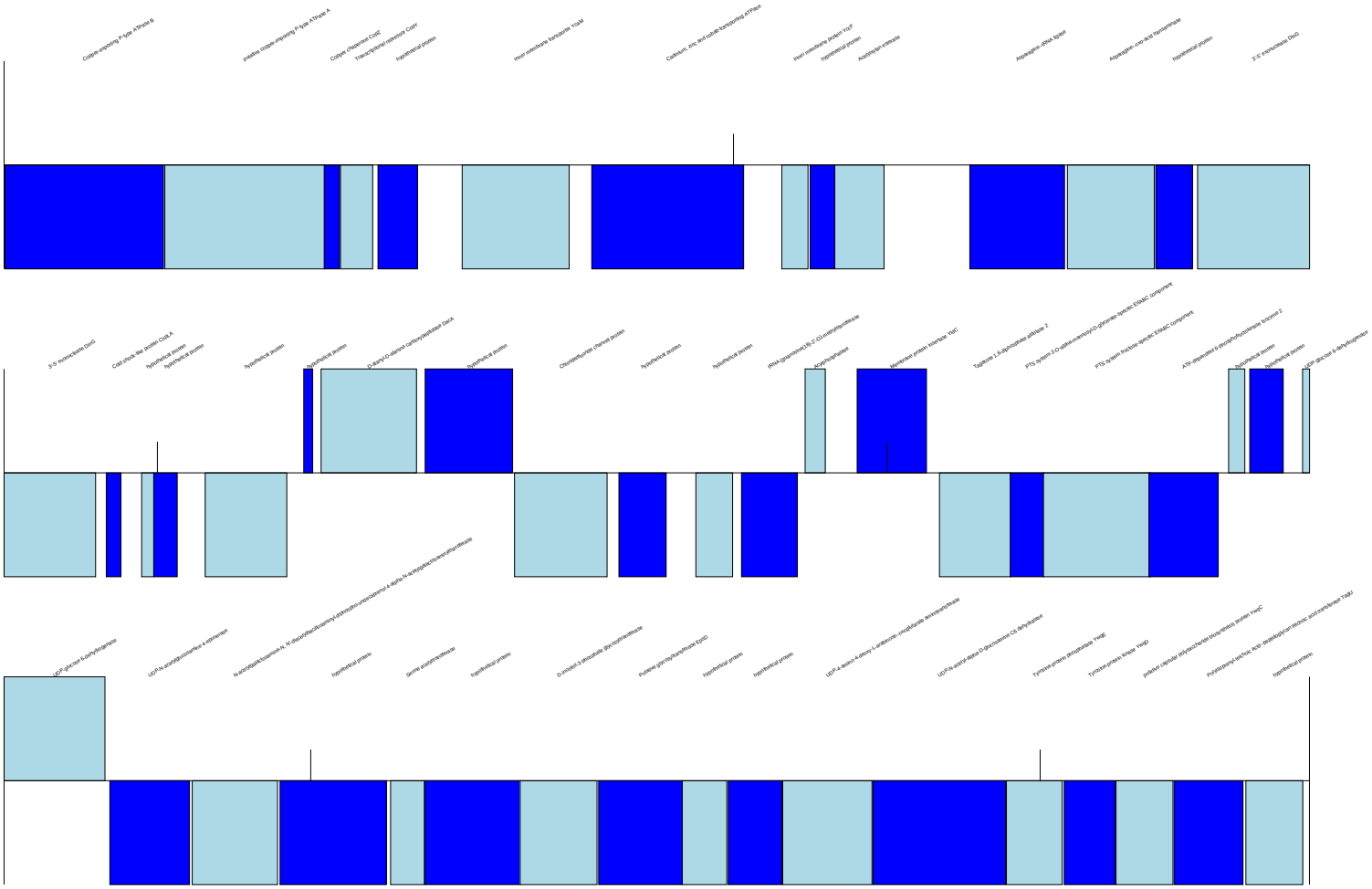


Figure 4(*d*): *arnB* contig in Sample 18. Positive sense genes are given above the line and negative sense are below the line. Genes are coloured in alternating blue for ease of visualisation

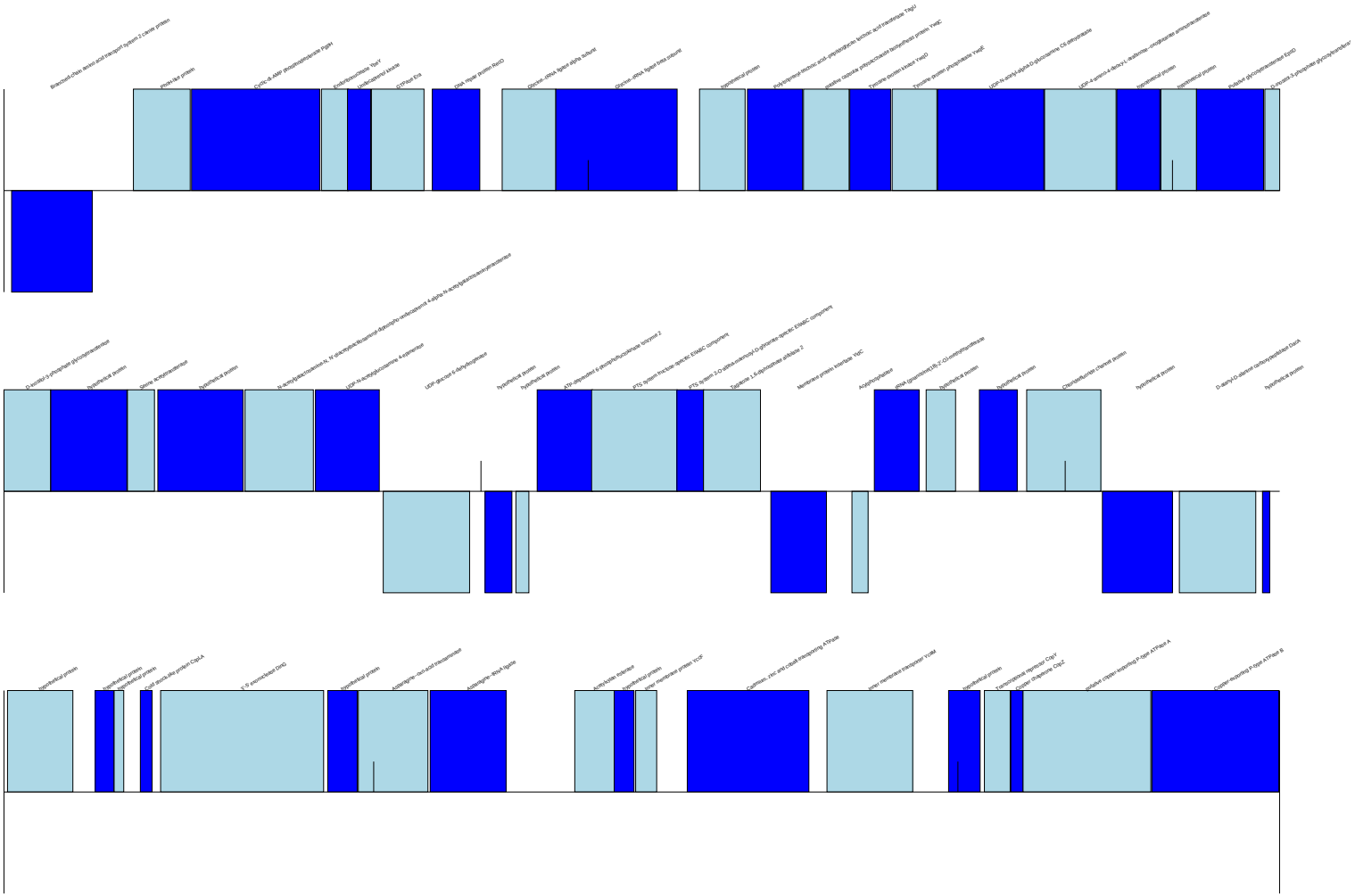


Figure 4(*e*): *arnB* contig in Sample 25. Positive sense genes are given above the line and negative sense are below the line. Genes are coloured in alternating blue for ease of visualisation