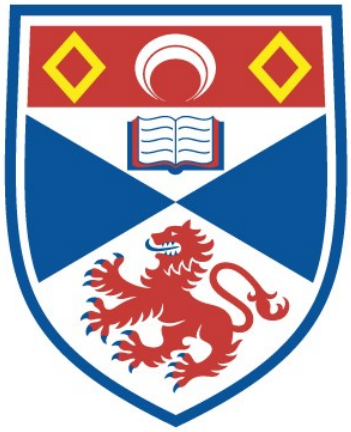
Genomic investigation of antimicrobial resistant enterococci

Martin Patrick McHugh

A thesis submitted for the degree of PhD
at the
University of St Andrews



2024

Thesis declaration

Candidate's declaration

I, Martin Patrick McHugh, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 35,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree. I confirm that any appendices included in my thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

I was admitted as a research student at the University of St Andrews in October 2015.

I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

Date 16/08/2023 Signature of candidate

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree. I confirm that any

appendices included in the thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

Date 16/08/2023 Signature of supervisor

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Martin Patrick McHugh, have obtained, or am in the process of obtaining, third-party copyright permissions that are required or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

No embargo on print copy.

Electronic copy

No embargo on electronic copy.

Date 16/08/2023 Signature of candidate

Date 16/08/2023 Signature of supervisor

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Martin Patrick McHugh, hereby certify that no requirements to deposit original research data or digital outputs apply to this thesis and that, where appropriate, secondary data used have been referenced in the full text of my thesis.

Date 16/08/2023 Signature of candidate

Thesis abstract

Enterococcus faecium and *Enterococcus faecalis* are important causes of healthcare-associated infections in immunocompromised patients. Enterococci thrive in modern healthcare settings, being able to resist killing by a range of antimicrobial agents, persist in the environment, and adapt to changing circumstances. In Scotland, rates of vancomycin resistant *E. faecium* (VREfm) have risen almost 150% in recent years leaving few treatment options and challenging healthcare delivery. Resistance to the last line agent linezolid has also been detected in *E. faecalis.* Whole genome sequencing (WGS) allows investigation of the population structure and transmission of microorganisms, and identification of antimicrobial resistance mechanisms. The aim of this thesis was to use WGS to understand the molecular epidemiology of antimicrobial resistant enterococci from human healthcare settings in Scotland. Analysis of some of the earliest identified Scottish linezolid-resistant *E. faecalis* showed the resistance mechanism, *optrA*, was present in unrelated lineages and in different genetic elements, suggesting multiple introductions from a larger reservoir. To inform transmission investigations, within-patient diversity of VREfm was explored showing ~30% of patients carried multiple lineages and identifying a within-patient diversity threshold for transmission studies. WGS was then applied to a large nosocomial outbreak of VREfm, highlighting a complex network of related variants across multiple wards. Having examined within-hospital transmission, the role of regional relationships was investigated which showed that VREfm in Scotland is driven by multiple clones transmitted within individual Health Boards with occasional spread between regions. The most common lineage in the national collection (ST203) was estimated to

have been present in Scotland since around 2005, highlighting its persistence in the face of increasing infection prevention and control measures. This thesis provides a starting point for genomic surveillance of enterococci in Scotland, and a basis for interventional studies aiming to reduce the burden of enterococcal infections.

General Acknowledgements

I can't thank my supervisors Matt Holden, Stephen Gillespie, and Kate Templeton enough for their enthusiasm, positivity, and support over a long period of time. I would also like to thank Prof David Harrison for supporting my request to undertake a PhD while continuing in my role at NHS Lothian. Many thanks to Prof Tom Evans and Prof Al Leonard for financial support from the Scottish Healthcare Associated Infections Prevention Institute. Thanks to Dr Elizabeth Dickson, Dr Nitish Khanna, Prof John Coia, Dr Benjamin Parcell, Dr Gabby Phillips, Dr Martin Connor, Dr Keith Morris, and Dr Robert Weir. At St Andrews, thanks to Kerry Pettigrew for generating sequence data and training on the MiSeq, and thanks to Catriona Harkins, Katarina Oravcova, Rebecca Mekler, and Arun Decano for discussions. I'm grateful to Ramon Fallon, Joseph Ward, and Peter Thorpe for bioinformatics training and for answering my very silly questions. Thanks also to Prof Julian Parkhill, Dr Simon Harris, and Prof Nicholas Thomson for providing access to computing resources at the Wellcome Sanger Institute.

Particular thanks to Rebecca Dewar, Juliet Kenicer, Lynne Renwick, and Jill Shepherd for keeping the NHS work going when I've been consumed by PhD.

I am very lucky to have parents and siblings who believe in me and gave me encouragement from a young age to pursue my interests, which has directly led to this thesis - thankyou. Thanks to the Drayes, Craigs, Dechants, and the Super Saturday Crew for fun and relaxing times.

Finally, thanks to Louise for her unending patience, support, and sacrifices, and Eadie for keeping a smile on my face. I'm very lucky to have you and I love you both very much.

Funding

This work was supported by the Chief Scientist Office (Scotland) [grant number SIRN/10]; the Wellcome Trust [grant numbers 105621/Z/14/Z, 206194]; and the BBSRC [grant number BB/S019669/1].

Research Data/Digital Outputs access statement

Sequence data generated in this project have been uploaded to public sequence repositories under the following accession numbers:

- PRJEB36950 (https://www.ebi.ac.uk/ena/data/view/PRJEB36950)
- PRJNA877253 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA877253)
- PRJNA997588 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA997588)
- PRJEB12513 (https://www.ebi.ac.uk/ena/data/view/PRJEB12513)
- PRJNA997587 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA997587)

Table of contents

List of figures

List of tables

List of abbreviations

AMR	antimicrobial resistance
AMRHAI	antimicrobial resistance and healthcare associated infections
AST	antimicrobial sensitivity test
ABC	ATP-binding cassette
bp	base pairs
BSAC	British Society for Antimicrobial Chemotherapy
BSI	bloodstream infection
CA	community associated
CARD	Comprehensive Antibiotic Resistance Database
CC	clonal complex
cgMLST	core genome multilocus sequence typing
CDS	coding sequence
CHL	chloramphenicol
CI	confidence interval
CRISPR	clustered regularly interspaced short palindromic repeats
DEL	deletion
DG	Dumfries and Galloway
EUCAST	European Committee on Antimicrobial Susceptibility Testing
FDA	US Food and Drug Administration
GP	general practice
HA	hospital associated
ID	identification
INS	insertion

IPC	infection prevention and control
IQR	interquartile range
IS	insertion sequence
LZD	linezolid
MALDI-TOF MS	matrix-assisted laser-desorption ionisation time-of-flight mass spectrometry
MCMC	Markov chain Monte Carlo
MDR	multidrug resistant
ME	major error
MGE	mobile genetic element
MIC	minimum inhibitory concentration
MLST	multilocus sequence typing
MLVA	multiple locus variable number tandem repeat analysis
MNP	multiple nucleotide polymorphism
MRSA	methicillin resistant *Staphylococcus aureus*
MT	MLVA type
MYA	million years ago
NA	not available
NCBI	National Center for Biotechnology Information
ND	not determined
NHS	National Health Service
ONT	Oxford Nanopore Technologies
PacBio	Pacific Biosciences
PBP	penicillin binding protein
PCR	polymerase chain reaction

PFGE	pulsed field gel electrophoresis
PTS	phosphotransferase system
QAC	quaternary ammonium compound
rRNA	ribosomal RNA
RIE	Royal Infirmary of Edinburgh
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SD	standard deviation
SKA	split *k*-mer analysis
SMRT	single molecule real time
SNP	single nucleotide polymorphism
ST	sequence type
UTI	urinary tract infection
VLKC	variable length *k*-mer cluster
VME	very major error
VR	vancomycin resistant
VRE	vancomycin-resistant enterococcus
VREfm	vancomycin-resistant *Enterococcus faecium*
VSE	vancomycin-sensitive *Enterococcus*
VSEfm	vancomycin-sensitive *Enterococcus faecium*
WGS	whole genome sequencing
ya	years ago

Chapter 1 Introduction

1.1 Clinical Burden of Enterococcal Disease

Enterococci are Gram-positive cocci bacteria carried in the gastrointestinal tract of mammals, reptiles, birds, and insects [1,2]. Enterococci were first described in 1899 simultaneously from England and France [3–5]. For much of the 20th Century enterococci were classified within the genus *Streptococcus*, being differentiated by possessing Lancefield group D antigen, growth at 45°C, hydrolysis of aesculin in the presence of 40% bile, and growth in 6.5% NaCl. Based on DNA hybridisation and ribosomal RNA (rRNA) sequencing, these bacteria were moved into the *Enterococcus* genus in 1984, and at time of writing 84 species have been designated within the genus [6]. Despite historically considered commensals, in recent decades enterococci have been increasingly identified as causes of human infection [7]. In England, the incidence of enterococcal bloodstream infection (BSI) has increased 65.8% from 9.6/100,000 population in 2012 to 15.9/100,000 population in 2021 [8].

Enterococci have been implicated in urinary tract infections (UTIs) and endocarditis in the community, as well as healthcare associated UTIs, BSIs, abdominal infections, and wound infections. Risk factors for enterococcal infection are gut carriage of the organisms, immunosuppressive conditions (malignancy, solid organ or bone marrow transplant, kidney disease, diabetes, rheumatoid arthritis, and advanced age), breaches in the gut barrier (mucositis, surgery, trauma), or iatrogenic factors (long hospital stay, antimicrobial

use, and indwelling medical devices) [9–12]. As these risk factors are generally only present in hospital inpatients enterococci have become important causes of nosocomial infection, and are one of the leading causes of healthcare associated infections [13–15]. Enterococci are responsible for an estimated 440,000 deaths per year globally, mainly due to BSIs and intra-abdominal infections [16]. Mortality rates in enterococcal BSIs are high, estimated at 23-47% reflecting the challenging patient group these infections are often encountered in [11,17]. As well as high mortality, enterococcal infections also significantly increase length of hospital stay and healthcare costs which complicates the delivery of modern medicine [9,18–20].

1.2 Virulence factors

Enterococci do not express the overt virulence factors of other pathogens such as *Staphylococcus aureus*, and their success as pathogens mainly relies on their ability to survive in the antimicrobial treated gut and in hospital environments. Virulence factors are more prevalent in *Enterococcus faecalis*, which may explain why this species was the leading cause of enterococcal infections until recent increases in *Enterococcus faecium* cases [8,21]. The main virulence factors allow attachment to the host, immune evasion, or biofilm formation (Table 1.1) [22,23].

Table 1.1 Virulence factors in enterococci

Type	Name	Mechanism	Pathogenic association	Species	Reference
Attachment	Ace	Collagen binding protein	Endocarditis	*E. faecalis*	[24]
Attachment	Acm	Collagen binding protein	General pathogenicity	*E. faecium*	[25]
Attachment	Ebp	pilus	Endocarditis, UTI	*E. faecalis*	[26]
Attachment	Aggregation substance	Surface protein	Endocarditis	*E. faecalis*	[27]
Attachment	Esp	Surface protein	UTI, endocarditis, biofilm	*E. faecalis* and *E. faecium*	[28]
Immune evasion	Cyl	Cytolysin	General pathogenicity	*E. faecalis*	[29]
Immune evasion	GelE	Metalloproteinase	Endocarditis, complement disruption	*E. faecalis*	[30]

General stress response	PTS	Carbohydrate utilisation	Colonisation, biofilm, endocarditis	*E. faecalis* and *E. faecium*	[31]

PTS, phosphotransferase system; UTI, urinary tract infection

1.3 Intrinsic antimicrobial resistance in enterococci

Enterococcal infections in humans are predominantly caused by two species, *E. faecalis* and *E. faecium*. Both species are intrinsically resistant to therapeutic doses of many commonly used antimicrobials [32–35]. The cell wall of enterococci inhibits the passage of aminoglycosides and eliminates their clinical use unless given alongside a cell wall active agent such as ampicillin or vancomycin [22]. The presence of low affinity penicillin binding protein (PBP) in *E. faecalis* (PBP4) and *E. faecium* (PBP5) confer intrinsic resistance to cephalosporins, flucloxacillin, aztreonam, and temocillin [36]. Enterococci are capable of absorbing folate from their environment which overcomes the therapeutic action of trimethoprim-sulphamethoxazole, although they can test sensitive *in vitro* in media lacking folate which can lead to confusion [37]. Fluoroquinolone resistance is now widespread in *E. faecalis* and *E. faecium* due to mutations in the DNA gyrase subunit gene *gyrA* and topoisomerase IV subunit gene *parC* [38,39]. *E. faecium* carries a chromosomal ATP-binding cassette (ABC) efflux pump, *msr(C)*, which confers low level resistance to macrolides, and the spread of the *erm* rRNA methylases has led to most clinical

enterococci being macrolide resistant [40–42]. Resistance to clindamycin and streptogramins is common in *E. faecalis* and is conferred by ABC-F ribosome protection encoded by *lsa(A)* [43]. Although not common causes of infections, *Enterococcus casseliflavus* and *Enterococcus gallinarum* are notable as being intrinsically resistant to vancomycin due to chromosomal carriage of the *vanC* gene [44].

Despite their multidrug resistant (MDR) phenotype, *E. faecalis* isolates usually remain sensitive to ampicillin, vancomycin, linezolid, and daptomycin. Acquired resistance in *E. faecium* is an increasing challenge.

1.4 Acquired antimicrobial resistance in enterococci

Enterococci have grown in importance as human pathogens over recent decades, driven by their ability to acquire antimicrobial resistance (AMR) mechanisms to new agents after introduction into clinical use (Figure 1.1).

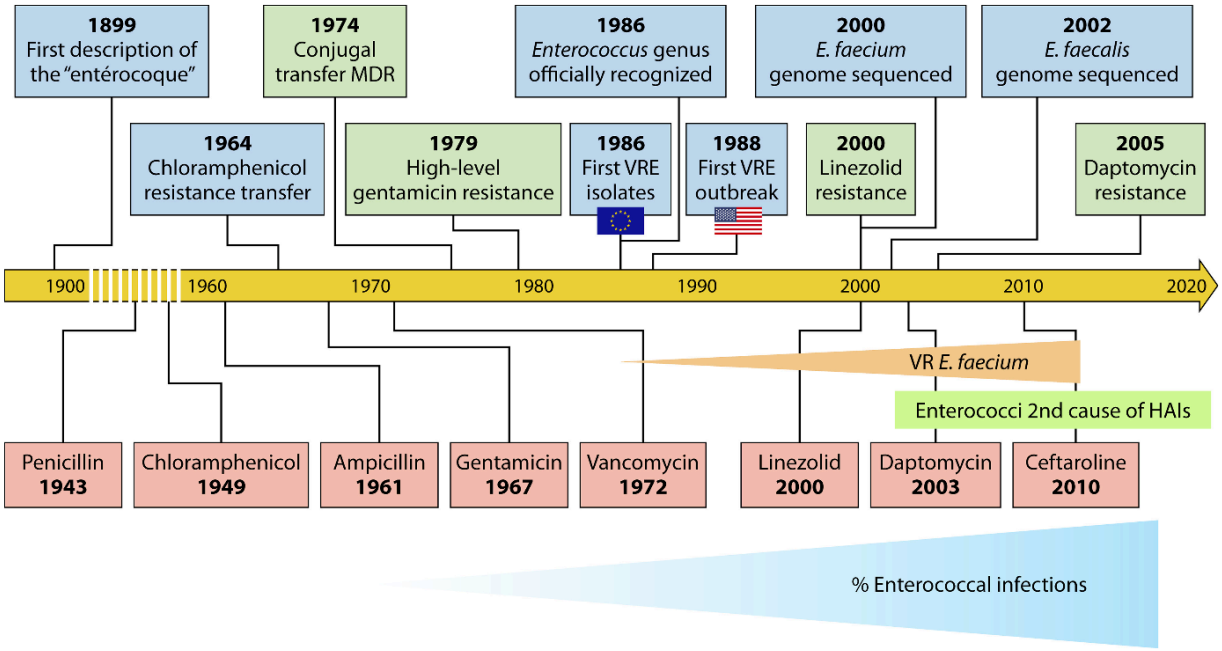


Figure 1.1 Timeline of enterococci as human pathogens

Relevant events are highlighted in blue rectangles, first detection of AMR in green rectangles, and introduction of antibiotics in red rectangles. MDR, multidrug resistant; VR, vancomycin-resistant. Reproduced with permission from Springer Nature [22].

1.4.1 Ampicillin

Ampicillin (and other β-lactams) block cell wall development by inhibiting crosslinking of peptidoglycan, leading to cell death. Ampicillin is the first choice treatment for enterococcal infections that test sensitive. Resistance to ampicillin is conferred by mutations in the *pbp4* or *pbp5* gene that eliminate drug binding [45,46]. This remains rare in *E. faecalis* (~2% of cases), but is widespread in *E. faecium* (>90% of cases) [8]. In *E. faecium*, the *pbp5* gene has been shown to be capable of mobilising from the chromosome, transferring to a recipient cell, and then inserting into the chromosome of the recipient by recombination which may explain the high prevalence of the resistant *pbp5* variant [47–49]. β-lactamase has been detected in *E. faecalis* but does not seem to be common [50].

1.4.2 Aminoglycosides

Aminoglycosides bind to the bacterial ribosome and impair proofreading of the mRNA template, leading to truncated or erroneous proteins which inhibits bacterial function. Despite their intrinsic low level resistance, enterococci can be treated with aminoglycosides in combination with a cell wall active β-lactam or glycopeptide. These combinations are recommended for severe enterococcal infections such as endocarditis [51]. Acquired resistance against aminoglycosides raises the minimum inhibitory concentration (MIC) to >512 mg/l, conferring high level resistance and eliminating their use even in combination therapy. High level resistance is conferred by three types of

aminoglycoside modifying enzymes that stop the drug being able to bind to the ribosome. *N*-Acetyltransferase AAC(6') acetylates an amino group of the aminoglycoside 2-deoxystreptamine nucleus, *O*-Adenyltransferases ANT(4'), ANT(6'), and ANT(9) adenylate a hydroxyl group, and *O*-Phosphotransferases APH(2'') and APH(3'') phosphorylate a hydroxyl group [52,53]. The main aminoglycoside used in enterococcal infections is gentamicin, resistance is widespread in *E. faecalis* and *E. faecium* being detected in >90% of tested isolates [54].

1.4.3 Glycopeptides

Vancomycin and teicoplanin are the main glycopeptide antibiotics, they inhibit cell wall synthesis by binding to D-ala-D-ala peptidoglycan sidechains and blocking their incorporation into the cell wall. A newer subclass are the lipoglycopeptides telavancin, oritavancin, and dalbavancin which show promising *in vitro* activity against enterococci but clinical experience is limited [35]. Resistance to glycopeptides is conferred by various *van* operon classes that modify the D-ala-D-ala side chain so that vancomycin can no longer bind (Table 1.2). The VanA operon is the most commonly encountered and contains seven genes - *vanS* encodes a transmembrane sensor, in the presence of glycopeptides it phosphorylates the operon repressor VanR which increases expression of the remaining five genes [55]. *vanH* encodes a dehydrogenase that reduces pyruvate to D-lactate, *vanA* encodes a ligase that binds D-ala to D-lac, the bound products are then added to peptidoglycan precursors by cellular machinery [22]. *vanX* encodes an amidase that cleaves

any D-ala-D-ala to further reduce glycopeptide binding, *vanY* encodes a carboxypeptidase which cleaves D-ala from cellular precursors, and *vanZ* encodes a gene of unknown function that contributes to teicoplanin resistance [56–58]. VanB is also encountered in clinical isolates but it does not confer teicoplanin resistance because $VanS_B$ does not recognise teicoplanin, although isolates have been known to become resistant on treatment due to modification of $vanR_B$ [59,60]. The other Van operons are not frequently identified.

Table 1.2 Vancomycin resistance mechanisms

Van Type	Vancomycin phenotype	Teicoplanin phenotype	Side chain modification	Presence in enterococci
VanA	R	R	D-ala-D-lac	Acquired
VanB	R	S	D-ala-D-lac	Acquired
VanC	R	S	D-ala-D-ser	Intrinsic
VanD	R	R	D-ala-D-lac	Acquired
VanE	R	S	D-ala-D-ser	Acquired
VanG	R	S	D-ala-D-ser	Acquired
VanL	R	S	D-ala-D-ser	Acquired
VanM	R	R	D-ala-D-lac	Acquired
VanN	R	S	D-ala-D-ser	Acquired

R, resistant; S, sensitive

Historically, the term glycopeptide-resistant enterococcus was used to refer to isolates carrying a *van* operon, more recently vancomycin-resistant *Enterococcus* (VRE) has become more accepted as this classification includes the *van* types that do not confer teicoplanin resistance (Table 1.2), VRE will be used for the rest of this thesis. VRE were

first reported in England and France in the 1980s and soon after were identified in the USA and other countries [61]. In the present day, <5% of *E. faecalis* are VRE while rates in *E. faecium* (VREfm) are higher and vary by country (Table 1.3). The quality of surveillance systems varies between countries, and in many countries no reliable data on VREfm rates are available. What information is available suggests vancomycin resistance is highest in the USA with Cyprus, Lithuania, and Brazil having similarly high rates, then Scotland, Australia, Eastern Europe, and the Balkans having VREfm rates of 30-50%. In Scotland, *E. faecium* BSI incidence has been relatively stable over 2008-2021, but the proportion that are VREfm has increased 148.2% (from 45 isolates in 2008 to 112 isolates in 2021, Figure 1.2). Rates have been rising across Europe and the UK, but it is currently unclear why the situation in Scotland is so profound [8,62].

Table 1.3 Reported VREfm rates around the World

Country	Type	Clinical syndrome	Period	VREfm (%)	Reference
USA	National surveillance	BSI	2018-2019	62.8	[63]
Brazil	Regional Surveillance	All	2007-2015	>60.0	[64]
Scotland	National surveillance	BSI	2021	40.4	[21]
Australia	National surveillance	BSI	2020	32.6	[65]
England	National surveillance	BSI	2021	21.0	[8]
India	Single Hospital	BSI	2020	19.2	[66]

Europe	International surveillance	BSI	2021	17.2 (range 0-66.4)	[62]
Africa	Metanalysis	All	2010-2019	10.2	[67]
China	National surveillance	All	2005-2017	<5.0	[68]

BSI, bloodstream infection

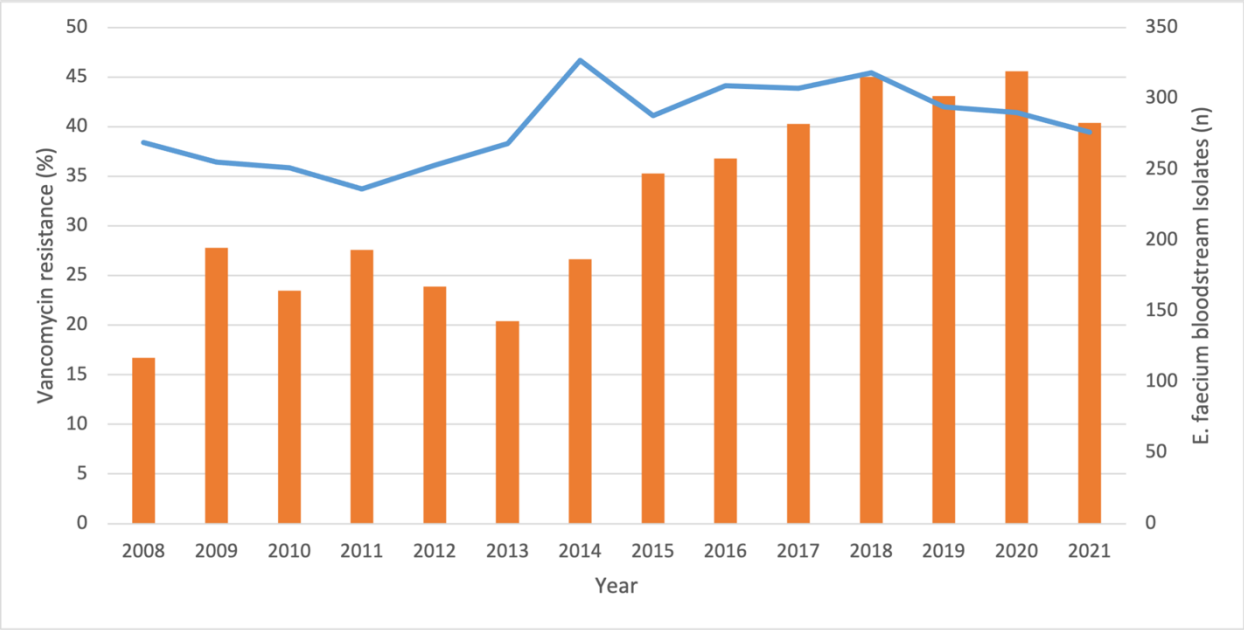


Figure 1.2 Rising vancomycin resistance in Scottish *E. faecium* BSIs

Number of *E. faecium* bloodstream isolates (line, plotted on right hand axis) and percentage vancomycin resistance (bars, plotted on left hand axis) in Scotland 2008-2021. Data collated from Scottish surveillance reports published by National Services Scotland **21,69**

1.4.4 Oxazolidinones

Linezolid and tedizolid belong to the oxazolidinone class of antimicrobials, they bind to the 50S ribosomal subunit and inhibit the formation of the 70S complex which then inhibits protein production by preventing the formation or elongation of peptide chains [70,71]. Linezolid has been available since 2001 and is a key antibiotic for the treatment of VRE infections, tedizolid has only been licensed since 2015 for skin and soft tissue infections and so there is less clinical experience against enterococci for tedizolid. Linezolid therapy is complicated by significant side effects (blood disorders, optic neuropathy, serotonin syndrome, among others) which limits recommended treatment duration to a maximum of 28 days [35]. Resistance is conferred by mutations in the 23S rRNA, usually G2576T or less commonly G2505A (*Escherichia coli* numbering) [72]. The 23S rRNA is present in six copies in *E. faecium* and four copies in *E. faecalis*, the number of mutated copies correlates with the MIC [73,74]. Mutation of the ribosomal proteins L3, L4, and L22 do not appear to be significant in enterococcal linezolid resistance [22]. Transferable linezolid resistance mechanisms have also been detected that can spread between cells via mobile genetic elements (MGEs). The *cfr*, *cfr(B)*, and *cfr(D)* 23S rRNA methyltransferases modify the ribosome to block binding of linezolid [75–77], and the *optrA* and *poxtA* genes are ABC-F ribosomal protection effectors that remove bound drug from the ribosome [78,79]. Linezolid resistance is generally rare, when encountered it is usually conferred by G2576T in *E. faecium* and *optrA* or *poxtA* in *E. faecalis*, although these associations are not fixed and all mechanisms have been detected in both species [80]. In Scotland, transferable *optrA*-

mediated resistance was first identified in 2015 in *E. faecalis*, although national surveillance shows linezolid resistance is currently at low levels (<2% in *E. faecalis* and *E. faecium*) [81].

1.4.5 Daptomycin

Daptomycin is a cyclic lipopeptide that inserts into the cell membrane in the presence of calcium, forming pores in the membrane which lead to ion leakage [22]. Daptomycin resistance is mediated by mutations in the *liaFSR* operon which encodes a regulatory pathway that responds to cell membrane stress. In *E. faecalis* resistance is due to movement of cell wall phospholipids away from the septum, in *E. faecium* there is a change in the charge of the membrane phospholipids leading to repulsion of daptomycin away from the cell [82–84]. Daptomycin resistance is rare, although MIC testing is only performed in select isolates as daptomycin is not included in commercial antimicrobial sensitivity tests (ASTs) systems such as the Vitek. Clinical use of daptomycin is further hampered because it is approved for complicated skin and soft tissue infections at a dose of 4 mg/kg body weight, or for *S. aureus* right sided endocarditis or BSI at 6 mg/kg [85]. Treatment of enterococcal BSI at these doses has been associated with treatment failure as the wild type MIC is around 2-4 mg/l [86,87]. Further studies have shown that doses of 10 or 12 mg/kg have suitable safety profiles and improve outcomes in enterococcal BSI [35,87]. Conversely, higher doses of daptomycin (>8 mg/kg) have been associated with muscle toxicity, and there has been a further association with eosinophilic pneumonia

necessitating weekly monitoring of creatine phosphokinase levels and perhaps also monitoring of daptomycin levels [88,89].

1.4.6 Tetracyclines

Tetracyclines inhibit protein synthesis by binding to the ribosome and blocking the A site, this in turn stops transfer RNA attachment in the P site. Tetracycline resistance is common in enterococci due to *tet(M)*, *tet(O),* or *tet(S)* mediated ribosomal protection, or *tet(K)* or *tet(L)* mediated efflux [35,90]. These mechanisms rule out clinical utility of older tetracycline derivatives such as doxycycline and minocycline. Newer derivatives are available that are active against these resistance mechanisms, tigecycline is the most well studied and is approved for use in complicated skin and soft tissue or intraabdominal infections. Tigecycline penetrates well into tissues but has low levels in serum which results in very poor outcomes when used alone for BSIs, where it must be used in combination therapy [91]. Resistance to tigecycline is currently very rare, but has been reported in *E. faecium* due to high expression of *tet(L)* or *tet(M)* [92].

1.5 Enterococcal Gut Carriage

Enterococci are carried in the gut often with no symptoms, however this can act as a reservoir for infection of other body sites if the patient develops one or more risk factors of enterococcal disease [93,94]. Gut carriage can also result in faecal-oral transmission

between individuals in healthcare settings, where the receipt of broad spectrum antibiotics reduces colonisation resistance against VRE in the gut [95–98]. Risk factors for VRE colonisation include exposure to antimicrobials (vancomycin, cephalosporins, and metronidazole, although others have been associated), frequent healthcare contact, prolonged hospitalisation, immunosuppression, intensive care admission, surgery, or indwelling catheters [35,99]. A systematic review over 8000 haematology patients identified 20% VRE colonisation rate, with a 24 times higher risk of BSI in colonised patients compared to uncolonized patients [99]. Similar VREfm carriage rates (19-24%) have been described in hospitals in the UK and Ireland [100,101]. In the general community population, gut carriage of VREfm appears to be uncommon (<1%) although this aspect is not well studied [100,102,103]. Healthcare workers are thought to be a potential source of patient infection, and VRE has been shown to survive on hands for up to an hour [104]. A study from the USA showed no VRE carriage in 755 healthcare workers suggesting limited staff-patient VRE transmission, although this the largest study of its kind to date it is limited by the single centre design [105].

The main factor promoting enterococcal colonisation of the gut is antimicrobial therapy [106]. Ampicillin, vancomycin, metronidazole, and neomycin have all been shown to reduce the diversity of the intestinal microbiota which then leads to an increase in *Enterococcus* either from within the existing microbiota or after acquisition from the hospital environment [107,108]. Once present, enterococci can adapt further via mutations or the acquisition of traits from other bacteria (see Section 1.8.1) that allows them to reach

dominance [106] (Figure 1.3). This intestinal dysbiosis can last for long periods of time after antibiotic, increasing the risk of host disease for months [107,109].

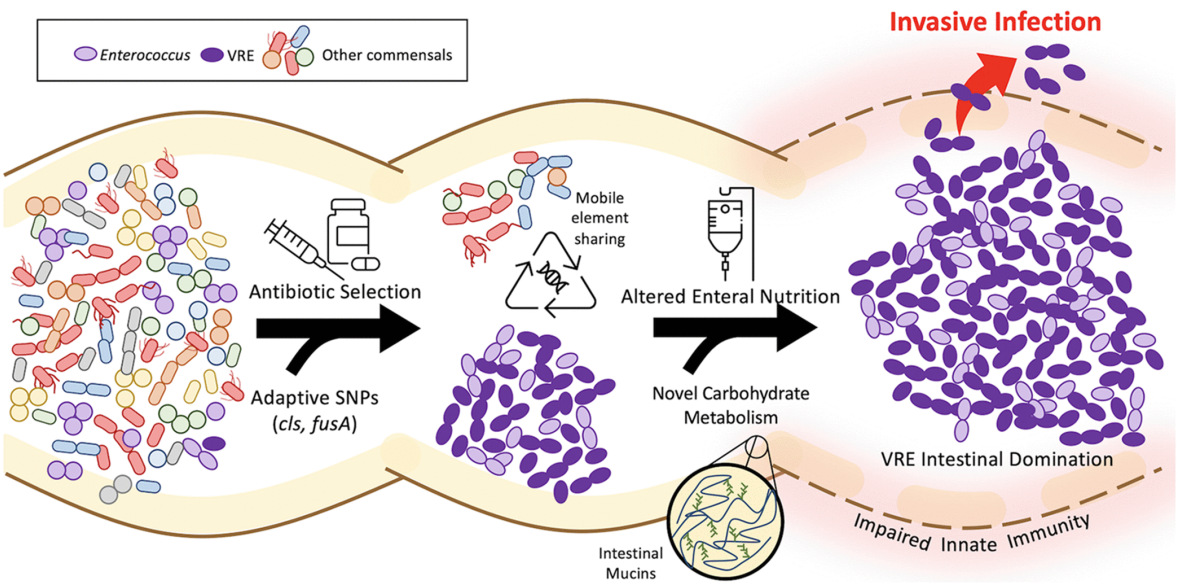


Figure 1.3 Adaption and dominance of enterococci in the antibiotic perturbed gut

Brown curved lines indicate the mucosal surface of the gut, yellow shading the mucus layer, black arrows external selective pressures, curved arrows genetic adaptations that allow enterococci (purple ovals) to resist eradication and outcompete commensal bacteria (multicoloured shapes). On the left is the normal gut microbiota, then the gut after antibiotic perturbation and enterococcal overgrowth, and on the right domination of enterococci and the breakdown of innate defences during time of ill health leading to invasive infection. SNPs, single nucleotide polymorphisms, VRE vancomycin-resistant enterococcus.

At the current time, there are no clinically validated decolonisation strategies to eliminate gut colonisation which would reduce the risk of transmission in healthcare and the burden of infections. Antibiotic-based decolonisation approaches tend not to be effective due to the broad spectrum of resistance in enterococci and their ability to acquire resistance to other agents [110]. Selective digestive decontamination involves the use of non-absorbed and broad spectrum antimicrobials to reduce microbial colonisation and infection in intensive care settings, the practice is popular in some European centres [111]. However, selective digestive decontamination studies have identified high incidence of enterococcal bacteraemia compared to non-intervention groups, likely due to the high use of cephalosporins and aminoglycosides in these regimes [112,113]. Faecal microbiota transplantation has had promising results in small studies evaluating the elimination of VRE carriage and reduction in subsequent clinical infections, but is not yet widely available in healthcare settings [114]. Regarding other potential strategies against enterococcal colonisation, no vaccine candidates are currently under clinical or preclinical evaluation, there has been some promising but early stage *in vivo* studies evaluating bacteriophage against enterococci, and bacteriocins show some potential for killing enterococci but are so far in preclinical development stages [115–118]. The duration of colonisation is also important, as if spontaneous decolonisation occurs then a patient can be cared for with standard precautions. However, the exact duration of carriage is not well defined and results from studies vary from nine weeks up to four years [119,120].

1.6 Control of Enterococcal Transmission in Hospitals

Enterococci can survive on inanimate objects for months to years [121,122], meaning that the ward environment can be a significant reservoir of enterococcal transmission [123–125]. Standard infection prevention and control (IPC) practice is to clean surfaces with detergent or chlorine based agents if known to be contaminated with VRE or body fluids, but outbreaks have been described even when standard practice is adhered to [126,127]. Novel decontamination methods are being developed, of which hydrogen peroxide vapour and ultraviolet light devices show promising *in vitro* reductions in pathogen survival and have begun to enter clinical use in some settings, although their widespread use are currently limited by high costs [128]. As well as cleaning of the environment and equipment, further measures to reduce transmission include eliminating transfer of bacteria between known carriers by use of contact precautions (gloves and fluid repellent gowns), isolation of carriers, and designation of care equipment as single patient use. Blane *et al*[129] identified VRE colonisation and infection rates halved in their patient population after moving to a new hospital with near 100% single occupancy rooms, environmental contamination also fell from 29% to 1-6%. This study highlights the interplay between the shared patient environment and transmission in hospitals, as well as the important role single occupancy rooms can have in limiting transmission of VRE.

Studies aimed at controlling VRE transmission show the need for multiple efforts including hand hygiene, patient screening, cohorting, isolation, environmental cleaning, and

antimicrobial stewardship [130–133]. Recently, two multi-year trials have investigated eliminating contact precautions for VRE in patient populations where VRE carriage was deemed to be endemic [134,135]. Multiple other measures were retained to reduce horizontal transmission. In both studies, incidence of VRE infection fell or remained stable during the study period suggesting elimination of contact precautions may represent a cost-effective strategy for management of VRE carriage. Up to 30 hospitals in the USA no longer use contact precautions for VRE, although an evidence review in 2015 could not identify high quality evidence to support or reject the use of contact precautions for endemic VRE [136]. Delivery of infection control services is currently challenging and the lack of international guidance on detection and managing VRE colonisation increases uncertainty.

1.7 Molecular typing methods

Bacterial typing is used to define genetic relatedness to establish relationships between isolates. Typing allows the study of population dynamics over time and space, particularly in the investigation of suspected transmission. Several typing methods have been developed for enterococci.

1.7.1 Multi Locus Sequence Typing

Multilocus sequence typing (MLST) is based on the sequencing of seven housekeeping genes [137,138]. Sequences are compared to a curated database, assigned to distinct alleles,

and the allele profile used to determine a sequence type (ST); closely related STs are grouped into clonal complexes (CCs). MLST targets conserved genes that diversify over decades, so is useful for comparing isolates over long periods of time and large geographic areas as opposed to local outbreak settings. The *E. faecium* MLST scheme has been available since 2002, recently it has been recognised that recombination within MLST loci can lead to some alleles being lost and so untypable with the current MLST scheme [139,140]. In 2023 Bezdicek *et al* [141] published an MLST scheme based on new loci which they found to have higher discrimination of STs than the previous scheme. An important factor of any typing scheme is consistency of comparisons over time so it remains to be seen whether the new scheme will be adopted by groups internationally.

1.7.2 Pulsed Field Gel Electrophoresis

Pulsed field gel electrophoresis (PFGE) is more discriminatory than MLST and has been used to investigate potential patient-to-patient transmission of enterococci. Genomic DNA is digested with a restriction enzyme, fragments are separated on a pulsed electrophoresis gel, and the banding pattern used to differentiate isolates [142]. PFGE was not widely adopted and remained a reference lab test due to turnaround time of five days or more, a lack of standard methodology, and the specialist equipment and staff required.

1.7.3 Multiple locus variable number tandem repeat analysis

Multiple locus variable number tandem repeat analysis (MLVA) amplifies various repetitive genetic regions, products are run on a gel with size markers and the band pattern used to determine an MLVA type (MT) [143,144]. MLVA is fast and relatively cheap, has higher resolution than MLST but lower resolution than PFGE [145]. MLVA was not widely used and there is no longer a curated database of MTs.

1.7.4 MALDI-TOF MS

In matrix-assisted laser-desorption ionisation time-of-flight mass spectrometry (MALDI-TOF MS) bacterial cells are crystallised into a matrix, ionised by a laser beam, the charged ions accelerated through a vacuum, and the particle time of flight measured [146]. A species-specific mass spectrum is generated and compared to a database to identify the bacterial isolate. This technique is rapid (around 90 seconds per isolate) and is widely used in diagnostic microbiology laboratories for identification of clinical isolates. Outside of species identification, MALDI-TOF MS has been used for rapid detection of VRE as well as outbreak analysis, with variable success [147–149]. Although the technology is available in many laboratories and this extra information would be clinically useful, currently MALDI-TOF MS is only used for species identification routinely.

1.7.5 Whole genome sequencing

Whole genome sequencing (WGS) provides information on the entire genome at single-base resolution, allowing pathogen identification, typing, and drug susceptibility in a single test [150]. WGS has high operational costs and the need for specialist staff and equipment to generate and analyse the data which has limited wide impact into patient care, although many reference centres provide WGS as part of their repertoire for public health and outbreak investigation. WGS can be performed with multiple platforms.

1.7.5.1 Illumina

Illumina sequencing platforms are the most commonly encountered instruments, they provide high quality data (PHRED score of 30 or 1 error per 1000 bases) and high throughput, with multiple instruments to cater for different size laboratories [151]. Read lengths are short, from 50-300 base pairs (bp) and multiple genomes can be sequenced per run, depending on the size of flowcell and instrument available. Run time on the sequencer is generally 1-2 days. For Illumina sequencing, DNA is prepared by shearing and ligating adapters and barcodes for identifying DNA from different samples, cleaned up and then loaded onto the sequencer. Illumina uses sequencing by synthesis technology where the ligated adapter is bound to probes into a solid phase flow cell and DNA polymerase elongates from the probe generating multiple clonal "clusters". Labelled nucleotides are sequentially added into the flowcell in the presence of DNA polymerase and the incorporation of nucleotides recorded for each cluster based on the emission of

fluorescence. Each flowcell contains millions of clusters all sequencing simultaneously to generate high throughput sequence data.

1.7.5.2 Oxford Nanopore Technologies

Oxford Nanopore Technologies (ONT) sequencing platforms are increasingly used for WGS due to their short run time, low purchase cost, and the ability to generate long reads up to millions of bases long [152]. The benefit of long reads is these can bridge repeats in the genome and provide a complete genome assembly, on the other hand short reads cannot resolve the genomic location of repeats longer than the read length and so the assembly is fragmented (Figure 1.4). ONT data has lower quality (PHRED 10-15, 1 error in 10-50 bases) due to random errors and systematic issues calling runs of the same base (homopolymers), however recent technology upgrades are bringing quality levels closer to Illumina [153,154]. For ONT sequencing DNA is prepared by ligating barcodes and adapters, and then on the sequencing flow cell the prepared DNA is fed through a protein nanopore embedded in a charged lipid membrane. As nucleotides pass through the membrane the charge is disrupted and the change in charge is interpreted to determine the sequence of each strand of DNA. ONT sequencing is real time in that data is immediately available to analyse as the sequencer is running, the user can decide to stop sequencing and wash the flowcell for further use or continue running to generate more data.

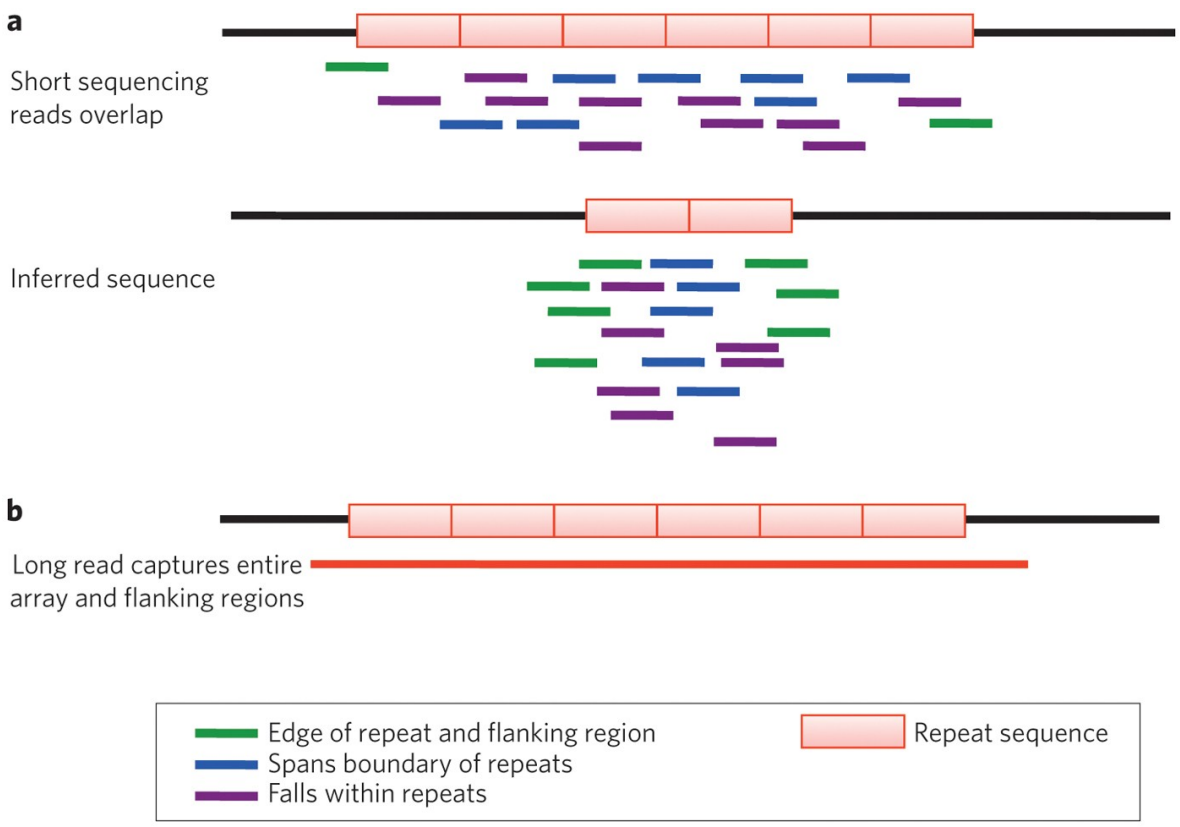


Figure 1.4 Long read sequencing allows assembly of repetitive genome sequences

Repeat elements (pink rectangles) in a genome of interest (black line) that are longer than the generated short read data cannot be resolved in genome assemblies, long reads overcome this limitation. A) Short reads that do not span an entire repeat can only provide information on the interface between two repeats (blue rectangle), the content of a repeat (purple rectangle), and the boundary with the rest of the genome (green rectangles), resulting in repeats being condensed in the inferred assembly but coverage being much higher than the rest of the genome. B) long reads can read through the repeat element, providing full genetic context and assembly of all copies. Reproduced with permission from Springer Nature[155]

1.7.5.3 Pacific Biosystems

Pacific Biosciences (PacBio) also provide long read sequence data but with higher accuracy than ONT and have become important platforms for the generation of high quality reference genomes [156]. The cost of PacBio instruments tends to be much higher than ONT which limits their widespread use. For PacBio sequencing double stranded target DNA is ligated at both ends with hairpin adaptors creating a single molecule real time (SMRT) bell. The SMRT bell is loaded into a SMRT cell containing nanoscale chambers which contain an immobilised DNA polymerase which binds to the SMRT bell. Fluorescently labelled bases are added and fluorescence measured as these are incorporated in real time, sequence generation continues as the closed SMRT bell continuously loops through the polymerase generating multiple passes of each base which increases the consensus accuracy of each read.

1.7.5.4 WGS data analysis

The common output of most WGS instruments are reads in FASTQ[157] format - these detail every sequence read for the sample as well as the quality of each base in the read. Read files can then be used in a range of tools to understand the genome of the input sample.

1.7.5.4.1 Reference-based mapping

A common approach is to map reads to a known high quality reference genome to identify how close the sample is to the reference, single nucleotide polymorphisms (SNPs) can be identified against the reference and used to infer the genetic distance. Commonly used software for reference mapping of short reads are BWA-MEM[158], Bowtie2[159], and SMALT (https://www.sanger.ac.uk/tool/smalt/), and for long reads Minimap2[160] is widely used. After mapping, reads are stored in the SAM/BAM format from which variants can be identified and a FASTA consensus sequence generated with tools such as SAMtools [161], GATK (https://gatk.broadinstitute.org/hc/en-us), and Freebayes[162]. For larger comparative studies, multiple consensus genomes based on mapping reads from different isolates to the same reference can be aligned together. Reference mapping often requires the masking of highly mutated regions such as MGEs and recombination which can confer multiple SNPs in a single genomic event, after masking the remaining regions are defined as the core genome. Recombination can be identified with Gubbins[163] or ClonalFrameML [164]. SNPs in the core genome are often utilised to build a phylogenetic tree to visualise the genetic relationships between query sequences using software such as RAxML[165], FastTree[166], IQ-TREE[167], or MrBayes [168].

1.7.5.4.2 *De novo* assembly

Reference based mapping only gives information on what is in the reference genome, when looking for novel genetic material a *de novo* assembly should be made based on read overlap within the sample of interest. Examples of common assemblers are Velvet

[169], SPAdes [170], and SKESA[171] for short reads, and Canu[172], Flye[173], Shasta[174], and Miniasm[160] for long reads. A growing application in genome assembly is hybrid assembly combining short and long reads from the same sample. Hybrid assembly combines long reads to bridge repeats, with short reads to provide low error rates to produce complete or near complete reference-quality genome assemblies [175]. Hybrid assembly can be performed short read first by generating a short read assembly then using long reads to bridge between contigs (e.g. Unicycler[176] uses this approach), or long read first by creating a long read assembly and then polishing with short reads to remove errors (Trycycler[177] uses this approach, but any long read assembler can be used in this way). Assembly polishing is an important step to improve the quality of an assembly based on long reads and remove potential indel errors [178]. Common polishing tools are Pilon[179] (short read only), Medaka (long read only, https://github.com/nanoporetech/medaka), Nanopolish (long read only, https://github.com/jts/nanopolish), and Racon[180] (either read type). Assemblies can be annotated with predicted coding sequences (CDSs) and other genomic features for further investigation with tools such as Prokka[181] or Bakta [182].

1.7.5.4.3 *K*-mer based approaches

Another approach to investigating genome contents is to split sequence data into blocks (mers) of size *k*, known as *k*-mers. *K*-mers can then be matched against a database to identify known characteristics (for example, match to known species for identification in Kraken2 [183]), or to other genomes to identify pairwise sequence matches [184]. *K*-mers can

be generated from raw reads or genome assemblies and are implemented in many popular *de novo* assemblers such as Velvet and SPAdes, and in the rapid long read mapper Minimap2. The advantage of *k*-mers is they are generally very fast to generate and compare, and they do not rely on a pre-defined reference so can utilise more variable regions of the genome [184]. Population partitioning using nucleotide *k*-mers (PopPUNK) utilises *k*-mers to determine the distance between sequences based on the core and accessory genome content. A model is fitted to the pairwise distances to cluster related genomes either with a Gaussian mixture model or hierarchical clustering [185]. Once a database of clusters is generated for a given species, new sequences can be added in without having to re-run the entire process which makes this attractive for ongoing genomic surveillance.

Standard *k*-mer approaches identify exact matches between the query sequences, and so cannot differentiate the presence of genetic variants such as SNPs from absence of the sequence in the genome. For this reason, *k*-mers are not usually able to distinguish relationships between related genomes (e.g. within an MLST sequence type). However, some tools are available for *k*-mer based SNP typing. kSNP utilises *k*-mers to identify SNPs between genomes by identifying the variant base in the middle of *k*-mers at variant loci [186]. A refinement of the *k*-mer approach is split *k*-mers, where a pair of *k*-mers have a gap (of 1 or more nucleotides) allowing the identification of conserved sequence surrounding variant regions. Split *k*-mer analysis (SKA) software has been shown to reliably and quickly cluster closely related genomes, such as those linked to recent transmission [187].

1.7.5.4.4 Options for further characterisation

Further characterisation of genome data is possible by comparing to databases of known markers. For example, MLST can be assigned by identifying the relevant alleles in the genome sequence under investigation either from reads using SRST2[188] or ARIBA[189], or from an assembly using ABRicate (https://github.com/tseemann/abricate). Other markers can be assigned such as AMR from the ResFinder[190], Comprehensive Antibiotic Resistance Database (CARD) [191], or AMRFinderPlus[192] databases, virulence from the VirulenceFinder[193] or VFDB[194] database, or plasmid replication genes from the PlasmidFinder[195] database. Assembled contigs can also be identified as likely chromosomal or plasmid using RFPlasmids[196] or mlplasmids[197].

1.7.6 Core Genome MLST

Core genome MLST (cgMLST) is an extension of the standard MLST process, but utilising WGS data to identify a species-wide core genome containing thousands of genes [198]. Gene loci are then identified in a genome assembly and compared gene-by-gene to the database of loci to determine the profile [199]. cgMLST schemes are stable typing methods that are publicly available, facilitating global collaboration and data sharing. However, cgMLST cannot resolve an allele differing by a single or multiple mutations, so cgMLST links are not as sensitive as core genome SNPs [184].

1.8 Genomic understanding of enterococcal population structure

Genomic data can be used to understand the evolutionary processes as the content of the genome shows how the bacteria respond to the environment. Enterococci are of particular interest and have been studied extensively over the recent past as they have become important nosocomial pathogens.

1.8.1 Mobile Genetic Elements

MGEs are small DNA molecules capable of transferring between bacterial cells. MGEs carry genes that allow niche adaptation, including AMR, environmental survival, virulence, and nutrient acquisition [200,201]. Enterococci are known to carry multiple MGEs, accounting for up to 30 % of the genetic content [202–205]. The high proportion of MGEs in enterococci has been linked to the loss of the clustered regularly interspaced short palindromic repeats (CRISPR)-Cas adaptive immune system, which otherwise recognises and removes foreign DNA from the cell [206]. This trade-off allows enterococci to diversify in the face of the harsh hospital environment but may also reduce protection from bacteriophage [207–209].

1.8.1.1 Bacteriophages

Bacteriophages are viruses that infect bacteria [210]. During phage genome packaging host DNA can inadvertently be packaged and then transferred to a recipient cell via phage transduction [211,212]. Phages sometimes carry beneficial genes and are retained within the bacterial chromosome; the integrated phages are termed prophages. Phage transduction of tetracycline and gentamicin resistance have been described and enterococcal isolates carrying phages have been shown to be more virulent *in vivo* [213,214].

1.8.1.2 Insertion sequences

Insertion sequences (IS) are very small DNA elements (700 bp – 2.5 kb) that code only for proteins involved in transposition of the element [215]. ISs are flanked by inverted repeats, allowing recombination with other DNA molecules and insertion. IS elements are widespread in hospital-associated enterococcal isolates [204,216]. IS*256* has been identified as an important driver of genome diversification in enterococci [217]. Genome wide IS*256* transposition was induced by activation of prophage, or antibiotic exposure, insertion of multiple IS*256* then drove diversification as a stress response.

1.8.1.3 Composite transposons

IS elements can form composite transposons when two copies insert on either side of resistance or pathogenesis genes. The flanking ISs allow mobilisation of the transposon to

other genome regions within the cell. Important examples in enterococci are the *vanB* positive Tn*1547* or gentamicin resistance transposon Tn*5281* [22].

1.8.1.4 Conjugative transposons

Conjugative transposons are short DNA elements (20 kb to 100 kb) flanked by IS elements that integrate into DNA molecules, can circularise, and spread between bacterial cells [218]. Conjugative transposons differ from composite transposons by encoding machinery for transposition as well as genes conferring AMR or other beneficial traits [219–221].

1.8.1.5 Plasmids

Plasmids are usually circular DNA molecules (~1 kb – 100 kb) that replicate independently of the chromosome. Plasmids can transfer to neighbouring cells via conjugation, the machinery for which is usually encoded by the plasmid as well as genes for replication, maintenance, antibiotic or heavy metal resistance, increased pathogenicity, or bacteriocins that inhibit competitor microbes [32,222,223]. The replication (*rep*) genes on plasmids are relatively conserved and can be used for typing purposes to identify the plasmid families present in individual isolates, up to ten plasmid families have been identified in enterococci [224]. Genomic analysis of plasmid sequences in over 1600 *E. faecium* identified that plasmid contents were key in determining the source of the isolate, showing that plasmids carry factors that mediate niche adaptation [225].

In *E. faecalis* pheromone responsive plasmids are common, cells carrying these plasmids are attracted to non-carriers that emit a pheromone molecule, on contact a mating pair is formed by the plasmid-encoded aggregation substance and plasmid transfer is highly efficient. pCF10 and pAD1 are clinically relevant examples that carry AMR genes or virulence factors, respectively [226,227]. Non-pheromone responsive plasmids are also encountered in *E. faecalis* and can have a broad host range, allowing transfer of material between species or genera [228]. *E. faecium* plasmids are not pheromone responsive, they often carry toxin-antitoxin systems to ensure plasmid survival (if the plasmid is lost, the toxin kills the cell) and multiple antibiotic resistance genes [22]. Recently, linear plasmids have been identified in *E. faecium* and other bacterial genomes using long read sequencing, these plasmids have palindromic repeats and structures to protect against degradation by genome defence mechanisms [229,230]. Linear plasmids have been identified to carry AMR markers including *van* loci and biosynthetic gene clusters allowing nutrient acquisition, but most identified CDSs are hypothetical so a lot remains to be uncovered about their function [230]. Often, plasmids in clinical isolates do not fit into only one of the descriptions above and display a hybrid structure due to multiple recombination events between plasmids and other MGEs [225].

1.8.2 Knowledge before the WGS era

Most understanding of enterococcal population structure up until the mid 2000s were based on MLST. MLST based studies showed the preponderance of specific STs causing infections in healthcare. These hospital associated (HA) lineages were enriched for genes encoding AMR, virulence, environmental survival, and often have reduced or lost CRISPR-Cas genome defence mechanisms [209,231–233]. Initially, the *E. faecium* HA lineage was defined as a single clonal complex (CC17) based on eBurst analysis [234], although eBurst was later shown to be inappropriate for analysing *E. faecium* populations due to the high rates of genomic recombination [235]. *E. faecalis* CC2 and CC9 were also defined as HA based on eBurst analysis [137], although many *E. faecalis* STs were found in hospitalised and healthy humans as well as animals, suggesting *E. faecalis* has evolved towards generalism and survival in multiple environments [233].

1.8.3 The WGS era begins

The study of microorganisms has been revolutionised in the past 20 years by the advent of WGS. Due to improvements in technology it is now possible to generate high-resolution WGS data in a few hours to days at relatively low cost [236]. At time of writing, the European Nucleotide Archive contains 31,162 *E. faecium* and 10,021 *E. faecalis* raw sequence datasets.

The first enterococcal genome was generated in 2000 from *E. faecium* strain DO, however the assembly was incomplete and was not analysed in the literature for some years [237].

The first genomic studies of enterococci focussed on *E. faecalis* with the complete genome of the first VRE to be isolated in the USA, strain V583, published in 2003 [205,238]. The chromosome of V583 was 3.2 Mb, with three plasmids sized 18-66 kb, *E. faecium* genomes are slightly smaller at 2.8 Mb. Analysis showed 25% of the V583 genome was made up of mobile or foreign DNA. Further studies comparing V583 to carriage and probiotic strains showed the high prevalence of MGEs was unique to V583, suggesting an important role of mobile DNA in pathogenic enterococci [239,240]. Draft genomes for other enterococcal species were published in 2010 with two complete *E. faecium* genomes being made available in 2012 [204,241,242]. Genomic comparisons of enterococci from different sources highlighted that MDR isolates were less likely to have functioning CRISPR-Cas systems compared to antibiotic sensitive strains and so have enlarged genomes carrying multiple MGEs [203,206,243,244]. As well as numerous MGEs, clinical isolates carried multiple AMR, colonisation, and virulence genes as well as functionally distinct carbohydrate metabolism, oxidative stress, and metal homeostasis pathways compared to community associated (CA) strains isolated as carriage populations in healthy humans [245]. Interestingly, although HA and CA *E. faecium* strains differ significantly at the genomic level, there are examples of recombination from CA to HA strains as well as recombination from outside *E. faecium* into the HA lineage [246]. Recombination hotspots were detected in carbohydrate metabolism and cell wall structural genes. These findings suggest niche-adaptation of specific strains to different environments, particularly in HA strains [247].

More recently, a large comparative genomics study of enterococci and related organisms has estimated the emergence of enterococci to 425-500 million years ago, coinciding with colonisation of land by aquatic organisms (Figure 1.5) [248]. The transition from aquatic hosts to land-dwelling hosts selected for survival to desiccation, starvation, and disinfection which preluded the dominance of *E. faecalis* and *E. faecium* in modern healthcare settings.

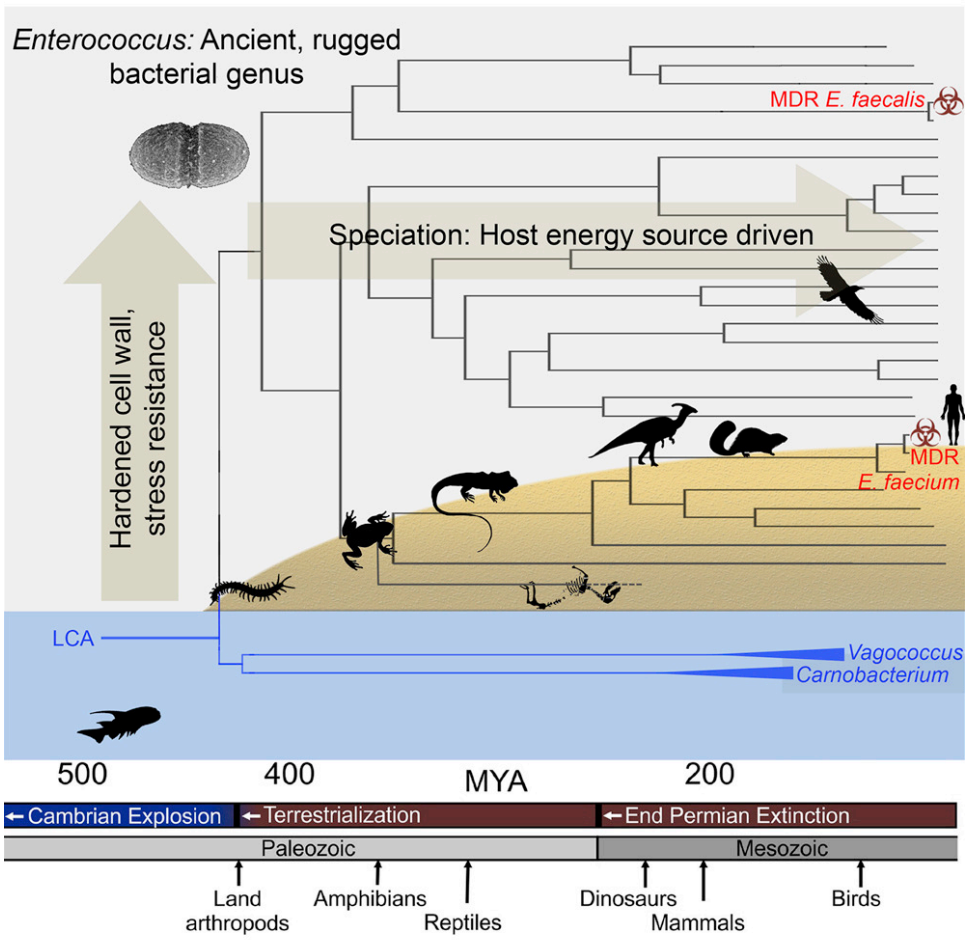


Figure 1.5 Evolutionary history of enterococci over millions of years

Diagram showing the ancestors of enterococci as commensals of aquatic animals, and the likely evolutionary origin of *Enterococcus* around the time of territorialisation, a period of adaptation to different land hosts and desiccation, and then a species explosion around the End Permian Extinction along with increasing speciation of land mammals. The traits required to survive these events led to organisms well adapted to survival in modern hospitals. MYA, million years ago. Reproduced with permission from Elsevier [248]

WGS can also determine the rate of genetic change by mutation over time when the sampling date is known for studied genomes. Mutation rates are useful as a marker of how quickly an organism can adapt to its surroundings and can be used as a molecular clock to determine evolutionary relationships between genomes. Estimated mutation rates in *E. faecalis* are 2.5-3.4 SNPs per year and in *E. faecium* 7 SNPs per year [249,250]. These mutation rates are not fixed and can vary between different genetic regions, being higher in recombination blocks and MGEs which are often masked from genomes prior to SNP comparisons. Different subclades of bacterial species can also have different mutation rates which can be an important consideration in SNP-based studies.

1.8.4 *E. faecium* population structure

The adaptation of specific *E. faecium* strains to the nosocomial environment has been further investigated using WGS. Studies comparing to MLST suggested rather than HA strains belonging to the single CC17, they were actually split into Bayesian analysis of populations groups 2-1 (ST78) and 3-3 (ST17, ST18), although the phylogenetic relationship between these groups remained largely unknown [251]. Lebreton *et al.* [209] performed phylogenetic analysis of 73 *E. faecium* from various isolation sources which demonstrated clear distinction between human carriage isolates (clade B) and isolates from animals and hospitalised humans (clade A), confirming previous findings based on MLST only (Figure 1.6) [251]. The clade split was estimated to have occurred around 3,000 years ago, coinciding with increased urbanisation, hygiene, and animal domestication. A

further split was seen within clade A between animal (A2) and nosocomial (A1) isolates, estimated at around 75 years ago and coinciding with the introduction of antimicrobials into healthcare. The prevailing ecological factors around the time of these bifurcations likely provided selective pressures which contributed to adaptation, demonstrated by the acquisition of new traits on MGEs and the loss of other niche-specific functions by genome decay [209].

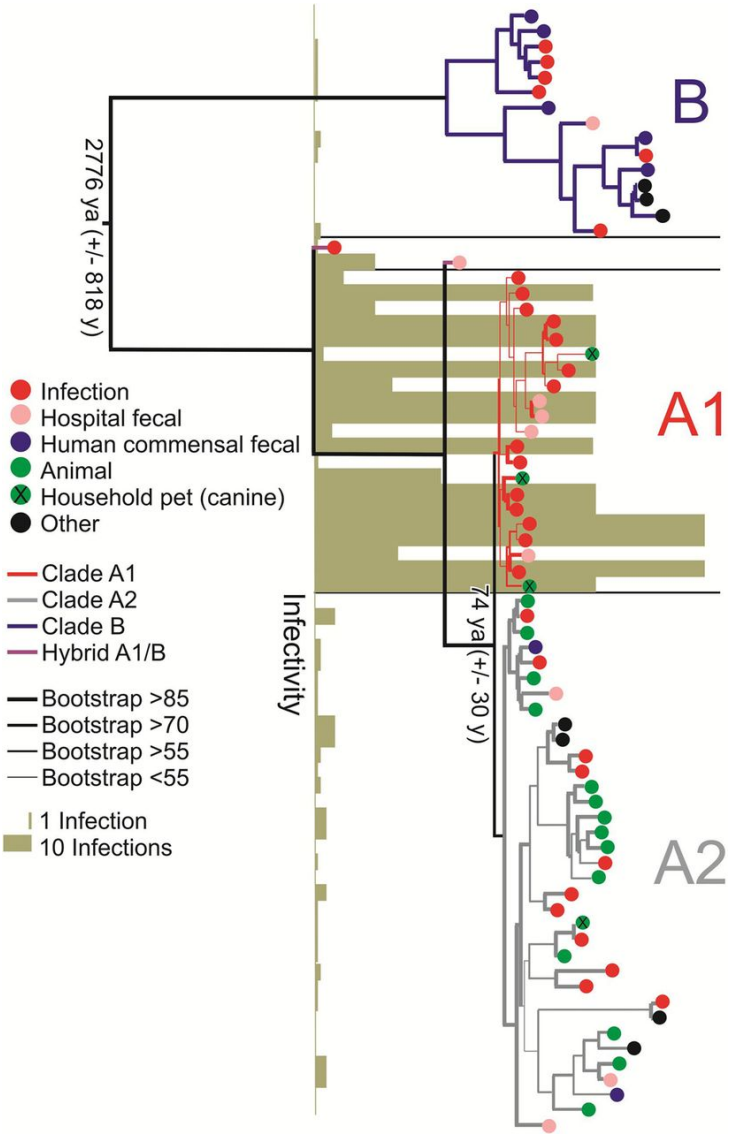


Figure 1.6 Phylogeny of *E. faecium* WGS data showing the split between Clade A1/A2 and B

Estimated dates printed on the branches (ya, years ago). The origins of isolates are indicated by coloured tips, the clade by coloured branches, and an infectivity score as gold bars. Phylogeny made in RAxML based on SNPs in core genes of 73 *E. faecium* genomes. © 2013 Lebreton *et al* [209], this figure, published in *mBio*, is available under a Creative

In 2016, further analysis of the population structure was performed by Raven *et al*[140], including 506 isolates from multiple host species they replicated the Clade A/B split previously identified by Lebreton *et al*[209] but did not identify differentiation within Clade A with regards to human and animal isolates (Figure 1.7). They suggested that nosocomial isolates rather represent a clonal expansion within Clade A from an animal-associated ancestor. This observation was supported by Arredondo-Alonso *et al*[225] who analysed a diverse collection of 1644 *E. faecium* isolates, focussing particularly on plasmid DNA. This analysis identified that nosocomial isolates had larger chromosomes as well as more and larger plasmids than carriage or animal isolates, and that plasmid content was more indicative than chromosomal content of the isolation source, again highlighting the role of MGEs in niche adaptation. Conversely, van Hal *et al*[252] studied 1128 *E. faecium* genomes and described the A1/A2 split as well as the presence of clade B genomes. This analysis identified that diversification within A1 was due to recombination with A2 and occasionally B strains, highlighting an important role for these non-pathogenic lineages to further drive A1 adaptation to the healthcare environment. The status of Clade A within *E. faecium* remains controversial, but there is clear evidence of differentiation between Clade A and Clade B, and recently Clade B strains have been formally reassigned into *E. lactis* [253].

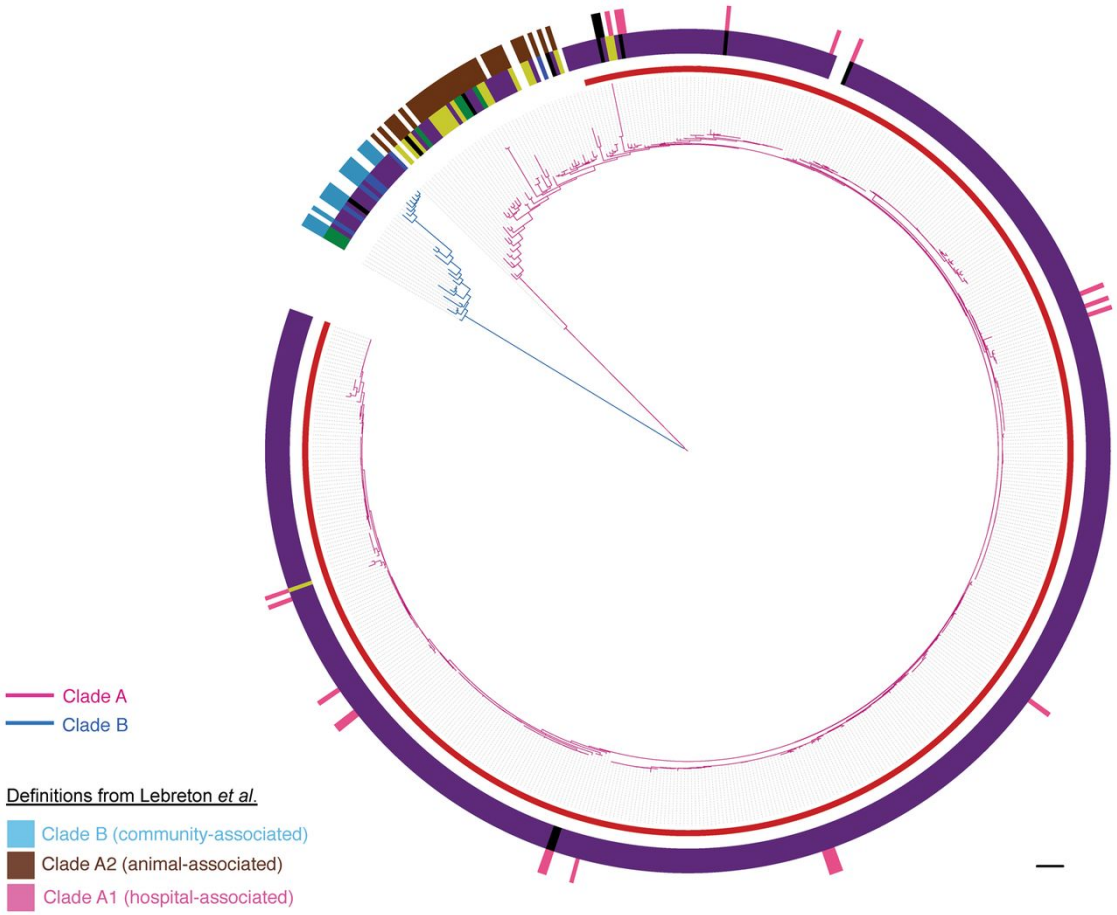


Figure 1.7 Genomic Population structure of *E. faecium*

Phylogeny based on SNPs in 1288 core genes from 579 isolates. Pink branches are Clade A, blue branches Clade B, the inner red ring indicates the clonal expansion of Clade A, middle ring the isolate source (yellow, animal; purple, clinical; blue, nonhospital; black, hospital; green, other; white, unknown), outer ring shows isolates from Lebreton *et al*[209]. Scale bar 9593 SNPs.

In recent years WGS has increasingly been applied to retrospectively analyse collections of *E. faecium* isolates from healthcare and understand their molecular epidemiology. Pinholt *et al*. [254] reported the first large enterococcal WGS molecular epidemiology study in which 132 VREfm from Danish hospitals in 2012-2013 were analysed. Isolates were classified into six groups based on core genome similarity, isolates in these groups were found in different geographical regions of Denmark as well as having closely related isolates from the same hospital and unit suggesting inter- and intraregional transmission. Further studies in Australia, Denmark, and the UK support the notion of sporadic interregional spread via patient transfers with ongoing intraregional transmission of local sub-clones within hospitals [140,249,255–258].

Molecular epidemiological analysis has not only been confined to VREfm; a number of studies have also analysed vancomycin-sensitive *E. faecium* (VSEfm) isolates. These show VSEfm and VREfm are commonly found intermingled within phylogenetic branches, suggesting that vancomycin resistance is variably gained and lost, particularly for *vanA* genes [140,249,255,258–261]. These findings suggest that control of VREfm may require infection control actions against vancomycin sensitive as well as resistant strains, while current practice would be to specifically target VRE only.

Some studies have compared WGS to other typing methods for *E. faecium*. All studies describe good overall agreement between WGS, MLST, and PFGE [254,262–264]. However, examples of isolates with the same PFGE type or MLST ST having hundreds or thousands

of SNPs difference in the core genome were seen, highlighting the higher resolution of WGS. These results show WGS is more discriminatory than existing methods for molecular epidemiology of *E. faecium*.

1.8.5 *E. faecalis* population structure

The molecular epidemiology of *E. faecalis* has been less well studied compared to *E. faecium*. Early analysis of 18 draft genomes by Palmer *et al*[244] found that the genomes showed little phylogenetic divergence and were closely related based on average nucleotide identity. Genomes did differ more in gene content with the pathogenic V584 sharing only 73% of its genes with commensals. This differed to *E. faecium* where pathogenic and commensal isolates were clearly different by average nucleotide identity. Raven *et al.* [250] analysed 168 UK and 347 global *E. faecalis* genomes and showed the presence of two internationally distributed lineages and one geographically restricted lineage enriched for AMR and virulence genes. Dating was successful for lineage one which suggested it had arisen in 1918 and undergone a clonal expansion in 1997. Despite the recognition of these lineages, 47% of clinical isolates did not fall within them highlighting the diversity within HA *E. faecalis* isolates compared to *E. faecium* which almost all fall into Clade A. Another study by Pöntinen *et al*[265] included 2027 isolates from a range of sources and countries spanning 1936-2018. No splits were identified in the population separating sampling sources suggesting strains were linked across different host species (Figure 1.8), this was also the case when plasmid contents were used for

clustering. Unlike in *E. faecium* there was no difference in chromosome of plasmid size between sampling sources. These studies support the hypothesis that *E. faecalis* is more of a generalist that can survive in many environments.

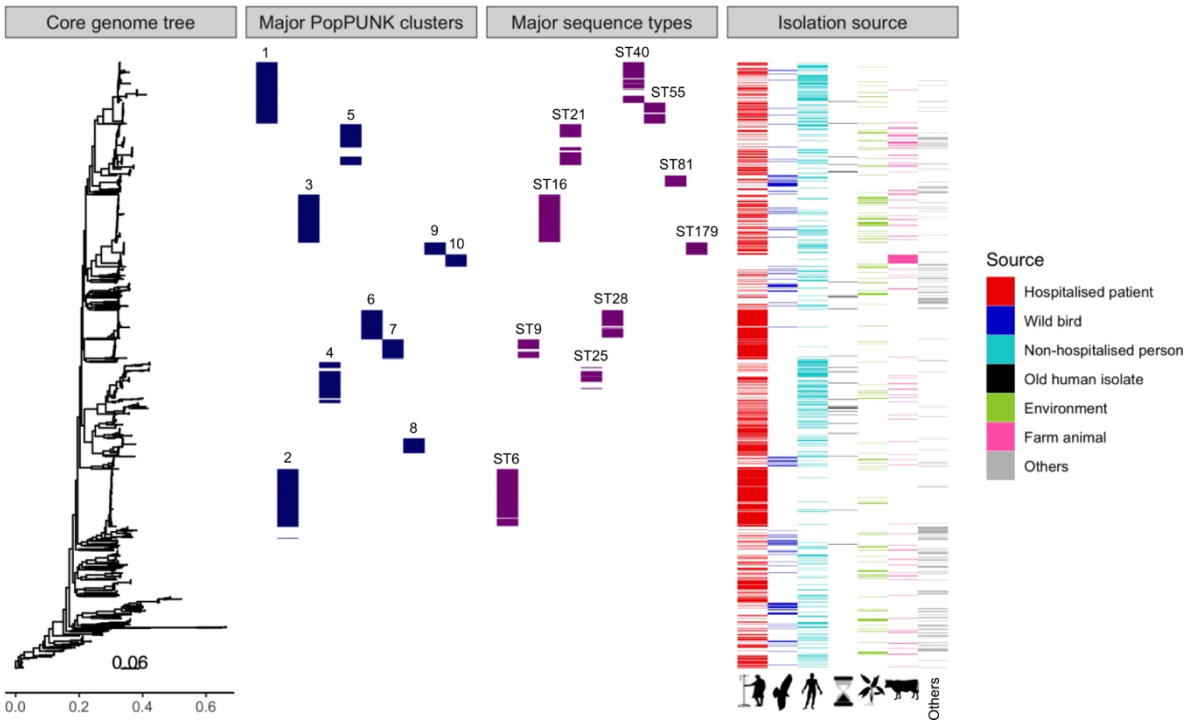


Figure 1.8 Population structure of *E. faecalis* shows interlinked lineages across different host types

E. faecalis (n=2027) from diverse collection sources and time periods, showing no clear clustering by isolation source with related genomes identified across ecological niches suggesting a generalist lifestyle. Maximum likelihood RaxML core genome tree in panel one based on mapping to the E07132 hybrid assembly (accession GCA_905123165). PopPUNK clusters are identified by blue blocks in panel two, MLST by purple blocks in panel 3, and isolate source in panel 4 (see figure for legend).

1.8.6 WGS for outbreak analysis

WGS has also been applied to understand hospital outbreaks of *E. faecium.* Lister *et al.* [266] describe an outbreak of *vanB*-positive VREfm colonisation on a neonatal ward involving 45 patients. They analysed 22 isolates by WGS showing they were all the same ST and highly clonal, suggesting transmission chains connecting patients and environmental reservoirs on the ward. Brodrick *et al.* [267] found 3/45 (7%) residents of a long-term care facility carried VREfm for up to 26 weeks, isolates between carriers were not related by WGS (>70 SNPs) ruling out direct transmission. Over time, carriers had 2-5 different *E. faecium* strains based on MLST, although the authors only analysed a maximum of two samples per week (one colony per sample) so were unable to determine the degree of co-carriage of different strains. Schlebusch *et al*. [147] investigate an outbreak of VREfm bacteraemia on a haematology unit where four patients had presented with BSI within one week. WGS of isolates from paired screening and bloodstream samples showed there were 3 distinct clones involved. Bashir *et al.* [268] used WGS to prove a VREfm infection in a liver transplant recipient came from the donor and was not a hospital acquired infection. The two isolates were MLST ST736 as were 3 other bloodstream isolates from the hospital, however WGS showed the two transplant-related isolates were almost identical from each other and formed a distinct cluster from other isolates. Finally, Raven *et al.* [249] analyse 293 *E. faecium* bloodstream isolates from their hospital. They identified six clusters of 93 isolates based on phylogenetic similarity and determined if any epidemiological links could be made for these clusters. They found evidence of patients sharing time on a ward for some

clusters, but others contained patients in different wards over long periods of time with no obvious links. These studies highlight that optimised WGS laboratory and bioinformatics protocols, but also high-quality epidemiological metadata are essential for investigating patient-level transmission.

1.9 Aims

Enterococci are characterised by broad AMR which cause complications in the management of immunosuppressed and comorbid individuals in healthcare. Vancomycin resistance is a particular challenge, and BSIs with VREfm have been increasing in Scotland in recent years, the reasons for this are unclear. Resistance mechanisms to reserved antimicrobials are also increasingly identified, while still uncommon it is key to understand the dynamics of these mechanisms in the enterococcal population to slow their spread into healthcare settings. WGS is now well accepted as the ultimate means of identifying the genetic basis of AMR, the evolutionary relationships between bacteria, and transmission dynamics between individuals. The central aim of this thesis was to use WGS to understand the molecular epidemiology of antimicrobial resistant enterococci from human healthcare settings within Scotland. This thesis provides a genomics-based understanding of the success of enterococci in Scotland, at multiple levels including individual isolates, individual patients, hospital wards, and regional and international relationships.

Genomics can provide unique understanding when the introduction of novel AMR markers is detected, it was hypothesised that novel AMR markers in enterococci would be present in closely related strains or plasmids. The aim of Chapter 3 was to identify whether newly identified *optrA*-positive *E. faecalis* in Scotland represent a clonal outbreak, spread of a plasmid, movement of a single resistance cassette, or multiple mechanisms. The objectives of the chapter were to use long and short read WGS to identify plasmids and examine the contents, as well as identify the genetic relationships between the generated bacterial chromosomes and plasmids.

After investigating a newly identified AMR threat, the next step was to understand how enterococci spread. As hospitals are important reservoirs for enterococci, and vancomycin resistance is very common, this was chosen as the setting. It was hypothesised that within-patient diversity could hamper the understanding of transmission patterns so the aim of Chapter 4 was to identify how diverse the VREfm population is in rectal carriage and determine the optimal number of colonies to use to effectively detect transmission. The objectives of this chapter were to identify diversity within colonised patients and infer how accounting for within-patient diversity affects transmission resolution.

It was hypothesised that nosocomial outbreaks would be caused by a single clone, and the findings from analysing within-patient diversity were then applied to an outbreak investigation. The aim of Chapter 5 was to investigate the utility of merged WGS and epidemiological analysis to understand suspected VREfm outbreaks. The objectives of this

chapter were to compare reference-free and core SNP-based clustering of genomes to each other and to PFGE, and to determine the linkage between patient epidemiology and genomic clusters.

After investigating the role of hospital-level transmission it was then hypothesised that different hospitals would harbour different enterococcal lineages. The aim of Chapter 6 was to identify the genetic background of VREfm disease isolates across Scotland. The objective of this chapter was to identify the strains present in different Scottish regions, identify AMR and plasmid markers, and determine the relatedness of Scottish and international VREfm genomes.

Chapter 2 General methods

2.1 Ethical Approval

Access to bacterial isolates as excess diagnostic material was approved by the National Health Service (NHS) Scotland BioRepository Network (Ref TR000126). This approval allowed access to minimal, non-identifiable patient metadata. Access to enhanced patient metadata was approved for NHS Lothian by the Caldicott Guardian (Ref 1690). This study was approved by the University of St Andrews Research Ethics Committee (Ref MD12651).

2.2 Data Availability

Sequence data generated in this project have been uploaded to public repositories under accession numbers PRJEB36950 (Chapter 3, https://www.ebi.ac.uk/ena/data/view/PRJEB36950), PRJNA877253 (Chapter 4, https://www.ncbi.nlm.nih.gov/bioproject/PRJNA877253), PRJNA997588 (Chapter 5, https://www.ncbi.nlm.nih.gov/bioproject/PRJNA997588), and PRJEB12513 and PRJNA997587 (Chapter 6, https://www.ebi.ac.uk/ena/data/view/PRJEB12513 and https://www.ncbi.nlm.nih.gov/bioproject/PRJNA997587).

2.3 Bacterial Isolates

Enterococci were isolated at participating institutions as part of routine clinical care and stored on Microbank beads (Pro-Lab Diagnostics, Wirral, UK) at -80°C until processed in this project. Isolates from the Royal Infirmary of Edinburgh (RIE) were isolated from CHROMID VRE agar (bioMérieux, Marcy-l'Étoile, France) for rectal carriage screening, or standard media for clinical samples, and identified by MALDI-TOF MS on a Microflex instrument (Bruker, Billerica, USA). Chapters 3 and 6 include isolates from other Health Boards where species identification was performed with MALDI-TOF MS on a Microflex or VITEK (bioMérieux) instrument, VITEK-2 GP card (bioMérieux), or API 20 Strep (bioMérieux). Antimicrobial sensitivity testing was performed in all centres with the VITEK-2 AST-607 card (bioMérieux) and interpreted with European Committee on Antimicrobial Susceptibility Testing (EUCAST) breakpoints [269]. For chapters 3 and 6, where multiple isolates were available from the same patient only the first isolate was included. For Chapter 4, three samples grew phenotypically distinct VREfm and both subtypes were stored by the routine laboratory, both subtypes were included here to compare to routine results. For Chapter 5, all growth on VREfm positive screening agar was removed and stored in a Microbank vial at -80°C, material was re-plated and 14 individual colonies selected at random per sample for DNA extraction.

For Chapter 3, linezolid resistance was confirmed by agar dilution at the Public Health England Antimicrobial Resistance and Healthcare Associated Infections (AMRHAI) reference laboratory and interpreted with EUCAST breakpoints [270]. Linezolid resistant isolates were then screened for the genetic determinant of resistance at the AMRHAI

reference laboratory. Detection of the G2576T mutation (*Escherichia coli* numbering) in the 23S rRNA genes was investigated by polymerase chain reaction (PCR) restriction fragment length polymorphism or by a real-time PCR-based allelic discrimination assay [271,272]. The *cfr* and *optrA* genes were sought by a multiplex PCR using primers for the detection of *cfr* (*cfr-fw*: 5'-TGA AGT ATA AAG CAG GTT GGG AGT CA-3' and *cfr-rev:* 5'-ACC ATA TAA TTG ACC ACA AGC AGC-3') [273] and *optrA* (*optrA-F*: 5'-GAC CGG TGT CCT CTT TGT CA-3' and *optrA-R*: 5'-TCA ATG GAG TTA CGA TCG CCT-3') (AMRHAI, unpublished).

For Chapter 5, PFGE of *Sma*I-digested DNA was performed at the Scottish methicillin-resistant *S. aureus* (MRSA) Reference Laboratory [274].

2.4 DNA Extraction

For DNA extraction performed at RIE (Chapters 3, 4, and 6), isolates were streaked onto Columbia blood agar (Oxoid, Cheshire, UK) or CHROMID VRE agar and incubated for 20-24 h at 37°C in air. A single colony was inoculated into 5 ml nutrient broth (E & O Laboratories, Falkirk, UK) and incubated for 20-24h at 37°C in air. Bacterial cells were pelleted (10 min, 5000 g), resuspended in 400 μl buffer P1 (Qiagen, Hilden, Germany), split into 200 μl aliquots, and one aliquot used for extraction while the other was stored at -20°C in case repeat extraction was required. Cells were disrupted first with lysozyme (20 μl of 100 mg/ml stock; Sigma Aldrich, St Louis, USA) for 1 h at 37°C, and then with

proteinase K (30 μl of 20 mg/ml stock; Qiagen) for 1 h at 56°C. Samples were cooled to room temperature and treated with ribonuclease A (4 μl of 100 mg/ml stock; Qiagen) for 30 min at 37°C. DNA was extracted from 200 μl of treated sample using the QiaSymphony® DSP DNA Mini Kit Version 1 (Qiagen) and the DNA Tissue protocol (Tissue_HC_200_V7_DSP) with DNA eluted into 100 μl buffer ATE (Qiagen).

For isolates extracted at the University of St Andrews by Dr Kerry Pettigrew (Chapters 3, 5, and 6), isolates were streaked to brain heart infusion agar (Oxoid, Basingstoke, UK) and incubated for 20-24h at 37°C in air. A single colony was inoculated into 5 ml brain heart infusion broth (Oxoid) and incubated overnight at 37°C in air. Two 1 ml aliquot were removed and pelleted (10 min at 10,000 g), one pellet was used for extraction and the other stored at -20°C in case repeat extraction was required. DNA was extracted from cell pellets using the Wizard Genomic DNA Purification Kit (Promega, Wisconsin, USA). Pellets were re-suspended in 480 μl 50 mM EDTA, 120 μl cell wall lysis mix added (containing 10 mg/ml lysozyme and 10 mg/ml lysostaphin (both Sigma Aldrich)), and then incubated at 37°C for 90 min. Lysates were centrifuged (16,000 g for 2 min) and supernatant discarded. Cell nuclei were lysed by adding 600 μl Nuclei Lysis solution and incubating for 5 min at 80°C. After cooling to room temperature 3 μl RNase was added and incubated at 37°C for 30 min. Protein precipitation solution (200 μl) was added, the mixture vortexed for 20 sec and incubated on ice for 5 min. Precipitated protein was pelleted (3 min at 16,000 g), the supernatant was transferred to a new 1.5 ml microtube containing 600 μl isopropanol, and mixed. Samples were then centrifuged (16,000 g for 2 min) and the supernatant

discarded. DNA pellets were washed with 600 μl 70% ethanol, centrifuged (16,000 g for 2 min), supernatant discarded, and any remaining ethanol was evaporated by incubating microtubes at 37°C for 30-45 min with the lid open. Cleaned DNA was rehydrated in 65 μl DNA Rehydration Solution at 37°C for 60 min or at 4°C overnight.

2.4.1 DNA Extraction Quality Control

The concentration of extracted DNA was determined using the Qubit dsDNA High Sensitivity Assay on a Qubit 3.0 Fluorometer (Invitrogen, California, USA). Purity of extracts was determined by the 260/280 nm absorbance ratio (acceptable range 1.65 – 1.95) after reading on a NanoDrop-1000 Spectrophotometer (Thermo Fisher, California, USA). RNA contamination and DNA integrity was checked by agarose gel electrophoresis. A 1% agarose gel was made with 100 ml Tris-EDTA and 10 μl SafeView Nucleic Acid Stain (NBS Biologicals, Cambridgeshire, UK). DNA extract (4 μl) was mixed with 1 μl BlueJuice Gel Loading Buffer (Invitrogen) and ran at 90 V for 40 min. To compare band size, 5 μl exACTGene 24 kb Max DNA Ladder (Fisher Scientific, California, USA) was included on every gel row. After electrophoresis, bands were visualised under ultraviolet illumination on a GelDoc XR with Quantity One 1D Analysis Software (BioRad, California, USA). A single band around 24 kb was indicative of intact genomic DNA with no RNA contamination. Extracts failing quality control checks were repeated. Negative controls were included on each extraction batch, these were accepted if no DNA was detected otherwise the extraction was repeated.

2.5 Whole Genome Sequencing

2.5.1 Illumina HiSeq Sequencing

This protocol was performed by the DNA Pipelines team at the Wellcome Trust Sanger Institute, Hinxton. Extracts were loaded into AFA tubes (Covaris, Brighton, UK) and sheared to 450 bp using a Covaris ultrasonicator. Short fragments were removed by solid phase reversible immobilisation bead clean-up on the Bravo Workstation (Agilent, California, USA) and library construction was performed with the Agilent SureSelect Kit. Libraries were barcoded and pooled prior to sequencing with a HiSeq 4000 (Illumina, California, USA) using 125 bp paired-end reads. Generated data was used in Chapter 6.

2.5.2 Illumina MiSeq Sequencing

Sequencing at the University of St Andrews was performed by Dr Kerry Pettigrew. Library preparation was performed with the Nextera XT kit (Illumina) which includes DNA fragmentation, tagging, PCR amplification, and clean up. Libraries were then pooled and sequenced with a MiSeq (Illumina) using 250 bp paired-end reads. Generated data was used in Chapters 3, 5, and 6.

Data for Chapter 4 was generated as above but with 600 bp paired-end reads, sequencing was split between the University of St Andrews (performed by Dr Kerry Pettigrew) and RIE (performed by the thesis candidate).

2.5.3 Oxford Nanopore Sequencing

Barcoded long read libraries were generated at University of St Andrews by Dr Kerry Pettigrew with the SQK-LSK108 1D Ligation Sequencing Kit (ONT, Oxford, UK) and sequenced with an R9.4 flow cell on a MinION sequencer (ONT). Basecalling and barcode de-multiplexing was performed with Albacore v2.1.3 (ONT) and the resulting fast5 files converted to fastq with Poretools v0.6.0 [275], or basecalled and de-multiplexed with Albacore v2.3.3 with direct fastq output. The generated data were used in Chapter 3.
For Chapter 4, long read libraries were generated at RIE by the thesis candidate with SQK-LSK109 Ligation Sequencing Kit (ONT) and sequenced with an R9.4.1 flowcell on a GridION sequencer (ONT). Live high accuracy basecalling and barcode dumultiplexing were performed in MinKNOW v19.12.6.

2.6 Sequence Analysis

2.6.1 Short read quality control

Fastq reads were trimmed with Trimmomatic [276] to remove low quality regions. Trimmomatic v0.36 and settings LEADING:5 TRAILING:5 SLIDINGWINDOW:4:15 MINLEN:100 were used for Chapter 3, to increase read lengths after trimming settings were modified for Chapter 4 (v0.32, MAXINFO:200:0.4) and Chapter 5 (v0.39, LEADING:3 TRAILING:3 MAXINFO:250:0.4 MINLEN:100). Data in Chapter 6 came from multiple sources with different read lengths, so the sliding window approach was used (v0.39 SLIDINGWINDOW:5:20 MINLEN:50). FastQC v0.11.9 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and MultiQC v1.12 [277] were used to identify adapter sequences in fastq files and these were also trimmed with Trimmomatic, if present (e.g. ILLUMINACLIP:TruSeq-PE-2.fa:2:30:10).
Quality trimmed reads were used in all subsequent analysis steps.

2.6.2 MLST

In silico MLST typing was performed using SRST2 v0.2.0 [188] and the relevant MLST database for the species (https://pubmlst.org/) sited at the University of Oxford [137,138,278,279]. When novel alleles were encountered, the mapping to each allele in the bam file was investigated in Artemis v18.0.0 [280]. Where the reads differed only by SNPs to the closest known allele a consensus was generated by repeating SRST2 with the --report_new_consensus flag and this was uploaded to PubMLST to assign a new allele. Where investigation of the bam showed uneven coverage (suggestive of mis-mapping by SRST2) or indels (not present in the SRST2 consensus due to inaccuracies from read

alignment) an assembly was produced with Unicycler v0.4.8 with default settings [176], the allele sequence was identified, and uploaded to PubMLST.

2.6.3 Short read reference-based mapping and phylogenetic trees

For Chapter 3, a collection of genomes from *E. faecalis* bacteraemia isolates were downloaded (BioProjects PRJEB4344, PRJEB4345, and PRJEB4346)[250] and quality trimmed short reads were mapped to the *E. faecalis* reference genome V583 (accession number AE016830) using the multiple_mappings_to_bam.py script (https://github.com/sanger-pathogens/bact-gen-scripts/blob/master/multiple_mappings_to_bam.py) with the SMALT v0.7.4 mapper (https://www.sanger.ac.uk/tool/smalt/). Mapped assemblies were concatenated and regions annotated as MGEs in the V583 genome (transposons, integrases, plasmids, phages, ISs, resolvases, and recombinases) were masked from the assembly by converting to Ns using remove_blocks_from_aln (https://github.com/sanger-pathogens/remove_blocks_from_aln). All sites in the alignment with SNPs were extracted using SNP-sites v2.4.0 [281] and pairwise SNP counts calculated (https://github.com/simonrharris/pairwise_difference_count).

For Chapter 4, reference based mapping was performed first to inform sample size considerations, and then to analyse genetic diversity of the study isolates. For sample size considerations, FASTQ data for 135 VREfm were downloaded as an example of within-patient diversity (BioProject PRJEB12937) [282]. Reads were mapped to the Aus0004 *E.*

faecium reference genome (accession number CP003351) with multiple_mappings_to_bam.py as above. As the objective was to differentiate population variants, to maximise the number of SNPs present recombination was not masked. For mapping of other genomes in Chapters 4-6, Snippy v4.6.0 (https://github.com/tseemann/snippy) was used with default settings and putative MGEs from the annotated reference genome (transposons, integrases, plasmids, phages, ISs, resolvases, and recombinases) were masked with the snippy-core command. Recombination was identified using Gubbins v2.4.1 [163], recombination blocks were masked from the core alignment using remove_blocks_from_aln (https://github.com/sanger-pathogens/remove_blocks_from_aln). All sites in the alignment with SNPs were extracted using SNP-sites v2.4.0 [281]. Pairwise SNP counts were calculated with pairwise_difference_count (https://github.com/simonrharris/pairwise_difference_count).

Hybrid assemblies generated from the collected isolates were used for mapping references in Chapter 4, Chapter 5 used the Aus0004 *E. faecium* reference genome (accession number CP003351), and Chapter 6 used Aus0004 for mapping the Scottish genomes and Aus0085 (accession number NC_021994) for mapping within variable length k-mer cluster (VLKC) 6_12_17_23_30 as this genome was phylogenetically closer to the VLKC members. Chapter 6 included data generated in this project and also a collection of short read datasets from the UK [250] and the rest of the World [283].

To compare Scottish ST1424 genomes in Chapter 6 a mixed approach was used as the isolates described by Lemonidis *et al*[284] only had assembled genomes publicly available. Genomes from Lanarkshire were downloaded from the National Center for Biotechnology Information (NCBI, BioProject PRJNA422798). The V24 complete ST80 reference genome (Accession CP036151)[285] was used as this lineage is phylogenetically close to ST1424 (Figure 6.3). ST1424 short reads from Chapter 4 were mapped to the reference genome and an alignment generated with Snippy v4.6.0. Lanarkshire draft assemblies were mapped to the reference with the nucmer option within ABACAS v1.3.1 (https://abacas.sourceforge.net/documentation.html) and unaligned contigs discarded. Mapped contigs were then added into the alignment with MAFFT v7.505[286] and the options --add-fragments and --keep-length to retain the original alignment size. The resulting alignment of all ST1424 was masked of MGEs identified in the V24 annotation as above, then passed to Gubbins v2.4.1 to mask recombination.

Phylogenetic trees were generated for Chapter 3 from the SNP alignment using RAxML v8.2.8[165] with 100 bootstrap replicates. For Chapters 4 and 5 the final SNP alignments were passed to IQ-Tree v2.0.3 with automatic model selection and 1000 ultrafast bootstraps [167,287,288]. For the national Scottish genomes in Chapter 6 (Figure 6.2) branch lengths were accurately reconstructed by passing the number of invariant sites to IQ-Tree using --fconst $(snp-sites -C <masked_alignment.fa>) and pairwise SNP distances inferred from the branch lengths in R v4.3.0 using the ape package [289,290]. Branch lengths were converted to substitutions per site by multiplying by the alignment size, then a distance

matrix constructed with the cophenetic.phylo function. For the comparison between Scottish national and cluster genomes, and within Scottish ST1424 in Chapter 6 (Figures 6.3 and 6.4) the RAxML v8.2.12[165] tree output by Gubbins v2.4.1 was used. All phylogenies were visualised with iTOL [291].

2.6.4 Hybrid *de novo* assembly

Hybrid assemblies were generated by combining trimmed short and long reads. For long reads adapters were trimmed and chimeric reads removed with Porechop v0.2.3 (https://github.com/rrwick/Porechop).

For Chapter 3, hybrid assemblies were generated with Unicycler v0.4.7 [176] in standard mode. Assemblies were annotated with Prokka v1.5.1 using a genus specific RefSeq database [292]. Hybrid assemblies were checked for indel errors using Ideel (https://github.com/mw55309/ideel) with the UniProtKB TrEMBL database v2019_1, and assembly statistics identified with assembly-stats v1.0.1 (https://github.com/sanger-pathogens/assembly-stats).

For Chapter 4, adapter-trimmed long reads were filtered to remove reads <1000 bp with Nanofilt v2.7.1 [293] and then assembled with the Trycycler v0.0.3 pipeline [177]. Briefly, reads were split into 12 total subsamples, three subsamples were each passed into four long read *de novo* assemblers (producing 12 assemblies in total): Flye v2.8.1, Redbean v2.5,

Raven v1.1.10, and Miniasm v0.1.3,[160,173,294,295]. A consensus assembly was generated within Trycycler and then polished with Medaka v0.11.5 (https://github.com/nanoporetech/medaka) and 2-3 cycles of Pilon v1.23 [179]. Assembly quality was assessed with assembly-stats v1.0.1 (https://github.com/sanger-pathogens/assembly-stats), Ideel (https://github.com/phiweger/ideel) with the UniProtKB TrEMBL database, and Busco v4.1.4 [296]. Polished assemblies were annotated with Prokka v1.14.6[181] using the Aus0004 reference genome (Accession CP003351) with the --proteins option. Abricate v1.0.1 (https://github.com/tseemann/abricate) was used to identify matches to ResFinder, VirulenceFinder, and PlasmidFinder databases[190,193,195,297], and putative prophages were identified with PHASTER [298]. Elements identified as plasmids had copy numbers estimated using short reads and Snippy: average depth for each plasmid was divided by the average depth of the chromosome. Nucleotide similarities between plasmids in the polished assemblies were estimated with Mash v2.2.2 [299].

2.6.5 Detection of AMR, plasmid, and virulence markers

AMR mechanisms, plasmid *rep* types, and virulence genes were detected using ARIBA (v2.12.1 for Chapter 3, otherwise v2.14.6) [189] and the ResFinder v3.0, PlasmidFinder v2.0.1, and VirulenceFinder v2.0.3 databases [190,193,195,297]. Resistance mutations against linezolid in the 23S rRNA (G2505A and G2576T based on *E. coli* numbering) [300], and fluoroquinolones in *gyrA* (S84R, S84I, S84N, S84L, S84Y, E88K, E88G, E88L, S98N) [39,301] and *parC* (S82R, S82I, E86A, E86K, E86T) [38,39] were also sought with ARIBA.

In Chapter 4, *de novo* assemblies were generated with short reads and Unicycler v0.4.8 using default settings [176]. AMR genes were then identified using Abricate v1.0.1 (https://github.com/tseemann/abricate) with default settings and the ResFinder 3.0 database [302]. Also in Chapter 4, to detect plasmids in isolates that had only undergone short read sequencing, plasmids from the two hybrid assemblies were used as references against all short read sets in Snippy and plasmids considered present if ≥85% bp were called with <20 SNPs/1000 bp.[285]

2.6.6 Comparison of phenotypic AST with *in silico* AMR detection

To compare the genotypic and phenotypic detection of AMR detection, a subset of 80 isolates from the national collection and the 87 isolates included in Chapter 5 were used. All isolates were from Lothian and had full Vitek AST results available, AMR was predicted with ARIBA as described in Section 2.6.5.

Resistance genes were deemed to confer resistance to different agents based on information in CARD [191]. For genotypic resistance, detection of 19 *pbp5* mutations (https://bitbucket.org/genomicepidemiology/pointfinder_db/src/master/enterococcus_faecium/phenotypes.txt) were considered to confer ampicillin resistance, detection of any *van* type was considered to confer vancomycin resistance and *vanA/D/M* considered to also confer teicoplanin resistance, detection of *aac(6')-Ie-aph(2'')-Ia* was considered to

confer high level gentamicin resistance, detection of *aac(6')-Ie-aph(2'')-Ia* or *aph(3')-IIIa* were considered to confer high level kanamycin resistance, detection of *ant(6)-Ia* was considered to confer high level streptomycin resistance, any detection of G2505A or G2576T (even at minority population variants) was considered to confer linezolid resistance, detection of *msrC* or any *erm* gene was considered to confer erythromycin resistance, detection of any *tet* gene was considered to confer tetracycline resistance, detection of *dfrG* was considered to confer trimethoprim resistance, and detection of *cat* was considered to confer chloramphenicol resistance.

Comparison was made between phenotypic AST and WGS with the phenotypic result being designated as the gold standard. Minor errors, major errors (ME) and very major errors (VME) were assigned based on US Food and Drug Administration (FDA) criteria [303]. Minor errors were only evaluated for trimethoprim as this was the only drug with intermediate range results.

2.6.7 Transmission Network Inference

In Chapter 4, all short reads were mapped to the VRED06-10 ST80 reference chromosome with Snippy, the V24 *E. faecium* ST80 genome (Accession CP036151) was included as an outgroup. An alignment of 2,814,943 bp was generated and 1,418,409 bp MGEs and recombination masked as described in Section 2.6.3. A posterior set of phylogenies were generated with MrBayes v3.2.7.[168] Two Markov chain Monte Carlo (MCMC) runs of four

coupled chains were run for 5,000,000 generations, sampling every 5000th. The final standard deviation of split frequencies was 0.013, the log-likelihood was stable, and the effective sample size of all parameters was >800, suggesting the model had converged. A random sample of 100 posterior trees was input to Phyloscanner v1.6.6.[304] Sankoff parsimony reconstruction was performed with *k* parameter of 281494.5, equivalent to a within-patient diversity threshold of 10 SNPs as used in other studies.[305] A transmission network was constructed in Cytoscape v3.9.0[306] showing edges with complex or transmission state and >0.5 probability. The role of smaller numbers of colony picks on transmission resolution was investigated by repeating the above with the first 3, 5, and 10 isolates randomly selected per sample.

2.6.8 PopPUNK Clustering

For input into PopPUNK in Chapter 5, short read assemblies were generated with SPAdes v3.15.5 using the --isolate flag and a minimum contig coverage of 15 [307]. PopPUNK v2.6.0 was then used with default settings to sketch a database of core and accessory distances for the 87 assemblies, and a Bayesian Gaussian mixture model was generated specifying four components [185]. The model gave seven clusters and a network score of 0.7923, refinement did not improve the model score, so the initial model was accepted. To identify close and putative transmission relationships between genomes, further subclustering was performed using PopPIPE v1.0.0 (https://github.com/bacpop/PopPIPE) with the minimum subcluster size set at 4.

For Chapter 6, a pre-built *E. faecium* database was utilised to generate VLKCs (https://www.bacpop.org/poppunk/). Short reads from Scottish, UK, and international collections were assigned to the database specifying a minimum *k*-mer count of 20 to remove low frequency *k*-mers introduced from sequencing errors.

2.6.9 Phylogenetic dating inference

The largest VLKC in Chapter 6 was investigated further by mapping to the phylogenetically related Aus0085 reference genome (accession NC_021996) using Snippy as described in Section 2.6.5. To identify node-defining SNPs, the MGE and recombination masked alignment and the final Gubbins tree were analysed with reconstruct_snps_on_tree.py (https://github.com/sanger-pathogens/bact-gen-scripts/blob/master/reconstruct_snps_on_tree.py). The final Gubbins tree and the year of sampling for each tip was used to infer node dates using BactDating v1.1 [308]. First, the optimal root of the phylogeny was inferred, then a root-to-tip analysis performed to roughly evaluate the temporal signal, and then the full MCMC Bayesian simulation ran for 10^5 generations using a relaxed clock to allow for variance in mutation rate across the phylogeny [309]. At the completion of the run all trace parameters were stable and the MCMC effective population size for mu/sigma/alpha were all >360, suggesting convergence. To further validate the temporal signal, the MCMC analysis was run again removing any effect of sampling date by setting all isolation dates to 2015 and the

deviance information criterion used to compare the models which confirmed the model run with true sampling dates was better giving further support to the temporal signal..

2.6.10 Detection of putative alcohol tolerance markers

Pidot *et al* [310] describe genetic markers in the ST796 *E. faecium* genome Ef_aus00233 (accession NZ_LT598663) associated with *in vitro* and *in vivo* tolerance of isopropanol, in Chapter 6 VREfm genome data from Scotland and elsewhere were searched for these markers (Table 2.1). Specific point mutations were identified in a sugar (glycoside-pentoside-hexuronide) symporter, RNA polymerase subunit, or in locations associated with a prophage. Additionally, the presence of ISEfa8 in association with a prophage, and a 70 kb region of a plasmid were also associated with tolerance. For the work presented here, point mutations were identified with Ariba and presence of specific regions with mapping. As ISEfa8 could be present in multiple genome locations, the complete prophage-ISEfa8 associated region in Ef_aus00233 was identified using PHASTER, which showed an intact prophage at 911595-958837 (which includes the ISEfa8 identified by Pidot *et al*) so this entire region was used for mapping to increase specificity. For detection with Ariba v2.14.6 [189], the complete CDSs of BN9748_RS02610 and BN9748_RS14440 were used and the identified amino acid changes searched for, the variants at position 2,396,698 bp and 2,397,781 bp were reported together in a prophage by Pidot *et al*, so in this study the region 2,396,690 to 2,397,790 bp was used with Ariba looking for mutations T9C (T2396698C) and G1092A (G2397781A). For mapping detection,

the regions indicated in Table 2.1 were used as reference sequences in Snippy v4.6.0 (https://github.com/tseemann/snippy) for short read mapping. The markers were considered present if ≥50% bases were called with <20 SNPs/1000 bp, the low coverage threshold used (≥50%) was to allow for variation in different genomes as the markers are either plasmid or phage associated, Pidot *et al* do not describe how much variation was present within the sequences they identified so a relaxed cut-off was used here to optimise sensitivity.

Table 2.1 Genetic markers in Ef_aus00233 associated with isopropanol tolerance[a]

Genome element	Tolerance Marker	Location	Locus tag	Product	Variant	Effect on tolerance	Detection method
Chromosome	Galactoside symporter	519608	BN9748_RS02610	Galactoside symporter	V264A	Increase	Ariba
Chromosome	RpoB	2838889	BN9748_RS14440	RpoB RNA polymerase subunit	H486N/Y	Decrease	Ariba
Chromosome	Prophage	2396698	Noncoding	Noncoding	T2396698C	Increase	Ariba
		2397781	BN9748_RS12235	Hypothetical protein	G2397781A		
Chromosome	ISEfa8	911595-958837[b]	Multiple	ISEfa8 + prophage	Presence	Increase	Mapping

Plasmid2	Plasmid	135000-9111	Multiple	Multiple CDS	Presence	Increase	Mapping

[a] as reported by Pidot *et al*[310]

[b] Pidot *et al*[310] report the ISEfa8 located at chromosome position 953094, in association with a prophage, in this study the entire prophage was identified and searched for CDS, coding sequences

A composite score was then generated for each genome based on the detection of the alcohol tolerance markers. The presence of the galactoside symporter variant, ISEfa8, or plasmid were each given a score of one, presence of either or both of the two prophage variants were given a score of one as the presence of both prophage variants was not identified as having additive effect on isopropanol tolerance [310], and the presence of either *rpoB* variant was given a score of minus one as these were found to decrease alcohol tolerance (Table 2.1).

2.6.11 Linear plasmid investigation

A hairpin structure was confirmed in the left hand side of p2_VRED06-10 with RNAstructure v6.0.1 [311]. The linear element p2_VRED06-10 was further investigated to determine if a circular element could be generated. PCR primers were designed using Primer-BLAST with the forward primer extending off the 5' end of the element, and the

reverse primer extending off the 3' end to generate a putative 205 bp product (p2_EXT_FOR, AGTCCCACGGAGGAAAAGAC; p2_EXT_REV TCTGTGGAGTGAAACAAAACC). A positive control was also designed to amplify a 390 bp region 500 bp upstream of the 5' end (p2_INT_FOR, ACCCAACGAAAAGGTTATCCAG; p2_INT_REV, TCGCTAACCCACACATACGG). PCR was performed on a ProFlex thermal cycler (ThermoFisher Scientific) with the Invitrogen *Taq* DNA Polymerase kit (ThermoFisher Scientific) - 5 µl 10x buffer, 5 µl dNTPs (2mM), 1.5 µl $MgCl_2$ (50mM), 2.5 µl primers (10 µM), 0.2 µl *Taq* polymerase, 33.8 µl nuclease free water, and 2 µl DNA extract. PCR was performed as follows: 94°C for 3 min; 30 cycles of 94°C for 30 sec, 58°C for 30 sec, 72°C for 60 sec; 72°C for 10 min; hold at 15°C. Products were ran on an Invitrogen E-Gel EX 1% Agarose gel (ThermoFisher Scientific) for 10 min with a 100 bp DNA Ladder (Promega, Southampton, UK), and visualised with the E-Gel Power Snap Electrophoresis System (ThermoFisher Scientific).

2.7 Data visualisations

Sequence visualisations and comparisons were generated with EasyFig v2.2.2 [312] or BRIG v0.95 [313]. Comparison of *optrA* sequences were performed with snipit (https://github.com/aineniamh/snipit). Epidemiological data were visualised with HAIviz v0.3 (https://haiviz.beatsonlab.com/).

2.8 Statistical Analysis

2.8.1 Sample size calculation

To determine the optimal number of colonies to analyse for within-patient diversity a power calculation was performed as described by Huebner et al [314]:

$$q \text{ ^ } n = (1 - P)$$

Where *q* = 1 – concentration of organisms, ^ = exponentiation operator, *n* = number of colonies sequenced, and *P* = probability of finding one or more variants.

2.8.2 Epidemiological support for genomic clusters

In Chapter 5, epidemiolocal linkage was established for all patient-pairs in the MLST, PFGE, core SNP, and PopPIPE clusters. Inpatient stay metadata were interrogated to identify patients on the same ward at the same time, patients on the same ward within 28 days of each other, patients on a different ward at the same time, and patients on a different ward within 28 days of each other. Any patient pairs who did not fit into these groupings were considered epidemiologically unlinked. The number and proportion of each epidemiological group were calculated for each clustering method, and 95% confidence intervals (CI) calculated.

2.8.3 Investigation of VREfm introduction and transmission on wards

In Chapter 5, to identify introductions and transmission of VREfm in the collection, patients were classified based on timing of positivity for VREfm. Patients positive ≤48 h of admission were considered already colonised and assigned as introductions, patients testing positive >48 h after admission having previously screened negative during their admission were considered as acquisitions during admission, and patients who tested positive >48 h after admission without previously testing negative were considered inconclusive. All PopPIPE clusters with confirmed introductions were further investigated to identify acquisition and inconclusive cases after introductions.

2.8.4 Software packages

R was used for statistical analysis [290]. In Chapter 4, presence/absence matrices of AMR genes were generated in R v4.0.5 using ggplot2 and patchwork packages [315,316]. In Chapter 6, Chi-square was used to identify significant associations between categorical data, relationships within the residuals were performed with the gplots and corrplot packages [317,318]. T-test was used to identify differences in the presence of putative alcohol tolerance markers between groups using the rstatix package [319].

Chapter 3 Presence of *optrA*-mediated linezolid resistance in multiple lineages and plasmids of *Enterococcus faecalis* revealed by long read sequencing

3.1 Introduction

Linezolid resistance is reported in ≤1% of bloodstream enterococcal isolates in the UK and is an important antimicrobial for the treatment of multi-drug resistant Gram-positive infections, including vancomycin-resistant enterococci [54,320]. The G2576T mutation in the chromosomal 23S rRNA genes can arise *de novo* during extended linezolid therapy [321], although antimicrobial stewardship and IPC measures appear to be successful in limiting the generation and spread of mutational linezolid resistance in clinical practice [322]. The methyltransferases Cfr, Cfr(B), and Cfr(D), and the ABC-F ribosomal protection proteins OptrA and PoxtA also confer resistance to linezolid in enterococci but are carried on MGEs, which can spread across genetically distinct lineages in the absence of antimicrobial selection [75,77–79,323–325]. Recent international surveillance confirms linezolid resistance remains rare, but *optrA* has recently spread to every continent and is the dominant mechanism of linezolid resistance in *E. faecalis*, despite first being identified as recently as 2015 [80]. Studies into the genetic context of *optrA* have identified the gene on both the chromosome and plasmids, often associated with insertion sequences such as IS*1216*, a possible vehicle for the rapid spread of *optrA* [326,327].

OptrA was first identified in Scotland in *E. faecalis* in 2016, a year after it was first described in China [78,81]. The aim of this chapter was to identify whether newly identified *optrA*-positive *E. faecalis* in Scotland represent a clonal outbreak, spread of a plasmid, movement of a single resistance cassette, or multiple mechanisms. It was hypothesised that the spread of *optrA* is driven by a single MGE, and to investigate this hybrid assemblies of short and long read sequencing data were made to generate complete genomes and to reconstruct the genetic environment of *optrA*.

Results presented in this chapter have been published in *Microbiology*:

https://doi.org/10.1099/mic.0.001137

3.2 Results

3.2.1 Detection of *optrA* in distinct *E. faecalis* strains

There were 14,133 isolates of *E. faecalis* in Forth Valley, Grampian, and Lothian between 2014 and 2017 (inclusive) from all sample types, 14 (0.1%) were identified as linezolid resistant, and eight (57.1%) of which were confirmed as *optrA*-positive at the AMRHAI reference laboratory. Six *optrA*-positive *E. faecalis* were available for further characterisation, these were among the first *optrA*-positive isolates identified from Scotland and so there was a public health and clinical interest in understanding their genetic background (Table 3.1) [81]. The earliest isolates in this collection were from the

Grampian region in the northeast of Scotland in 2014, 2015, and 2016. Three more isolates were identified in 2017 from the Lothian and Forth Valley regions in east and central Scotland (Table 3.1), with no clear epidemiological links between the patients. Only one patient had known exposure to linezolid prior to the isolation of an *optrA*-positive *E. faecalis*, two patients were hospitalised at the time of sample collection while the remaining four were from general practice. Samples were collected for symptomatic UTI or orchitis.

Table 3.1 Details of the *optrA*-positive *E. faecalis* characterized in this study

Isolate	Year	Region	Clinical Sample	Patient Source	MLST	Acquired linezolid resistance genes					Mutations in 23S rRNA		MIC (mg/l)	
						cfr	*cfr*(B)	*cfr*(D)	*optrA*	*poxtA*	G2505A	G2576T	CHL	LZD
WE0851	2014	Grampian	Urine	GP	480	-	-	-	+	-	-	-	≥64	8
WE0254	2015	Grampian	Urine	GP	19	-	-	-	+	-	-	-	≥64	8
WE0438	2016	Grampian	Urine	Hospital	330	-	-	-	+	-	-	-	≥64	8
TM6294	2017	Forth Valley	Urine	Hospital	585	-	-	-	+	-	-	-	≥64	8
BX5936	2017	Lothian	Semen	GP	894	-	-	-	+	-	-	-	≥64	8
BX8117	2017	Lothian	Urine	GP	16	-	-	+	+	-	-	-	≥64	8

CHL, chloramphenicol; GP, general practice; LZD, linezolid; MIC, minimum inhibitory concentration

WGS was performed to investigate the genetic relationship between the isolates and the context of the resistance mechanism. *In silico* MLST showed the six isolates belonged to different STs, suggesting they were genetically distinct (Table 3.1). To further confirm this, SNPs in the core genomes of the *optrA*-positive isolates were analysed which showed the isolates differed by a median 18,806 SNPs (range 13,909 – 22,272). Previous estimates suggest a genetic diversification rate of 2.5-3.4 SNPs/year for *E. faecalis*, highlighting the *optrA*-positive isolates share a very distant common ancestor [250].

3.2.2 *optrA* is carried on diverse MGEs

Hybrid assembly produced complete or near-complete genomes with <3% putative CDSs shorter than the closest reference match. This indicated the hybrid assembly process removed most indel errors, with 1-5% of CDSs expected to represent true truncated pseudogenes [328]. The hybrid assemblies contained between one and three plasmids ranging in size from 11-80 kb, with *optrA* present on a single complete plasmid in each isolate (pBX5936-1, pBX8117-2, pTM6294-2, pWE0254-1, pWE0438, pWE0851-1; Appendix 1).

The *optrA*-positive plasmids shared limited sequence similarity to the first described *optrA* plasmid (pE394, accession KP399637), with only the 5-10 kb region surrounding *optrA* and *fexA* (a chloramphenicol/florfenicol exporter) showing >70% nucleotide identity (Figure 3.1). In all six Scottish *optrA*-positive plasmids *optrA* and *fexA* were located within 550-750

nucleotides of each other intervened by a single CDS (hypothetical function in all but pBX8117-2 which was annotated as a putative NADH reductase). Within the Scottish *optrA*-positive isolates, plasmids pBX5936-1 (69 kb) and pTM6294-2 (53 kb) were most similar, sharing 97% average nucleotide identity over 40 kb aligned sequence (Figure 3.2). pTM6294-2 shared 99.8% identity with a 53 kb *optrA*-positive pheromone responsive plasmid detected in *E. faecalis* from a clinical sample in China (pEF10748), clinical samples in Spain (IsoBar1, IsoBar2, and IsoBar3), and raw dog food in Portugal (pAPT110) [329,330]. pWE0438 shared 92.3% nucleotide identity over 52 kb with pS7316 from an *E. faecalis* isolated from a hospitalised patient in Japan [331]. In pWE0438, the *optrA* and *fexA* genes were ~3.8 kb upstream of Tn*917* carrying *ermB*, and ~1.8 kb downstream of another Tn*3*-family transposase (Figure 3.2). pBX8117-2 carried *optrA* and the novel *cfr*(D) gene (encoding a 23S rRNA methylase that confers phenicol, oxazolidinone, pleuromutilin, and streptogramin A resistance) but apart from these genes showed no similarity to another *E. faecium optrA*/*cfr*(D)-positive plasmid identified in a clinical sample in Ireland (M17-0314) [332]. The other Scottish *optrA*-positive plasmids showed limited similarity to other published examples outside of the *optrA*/*fexA* region.

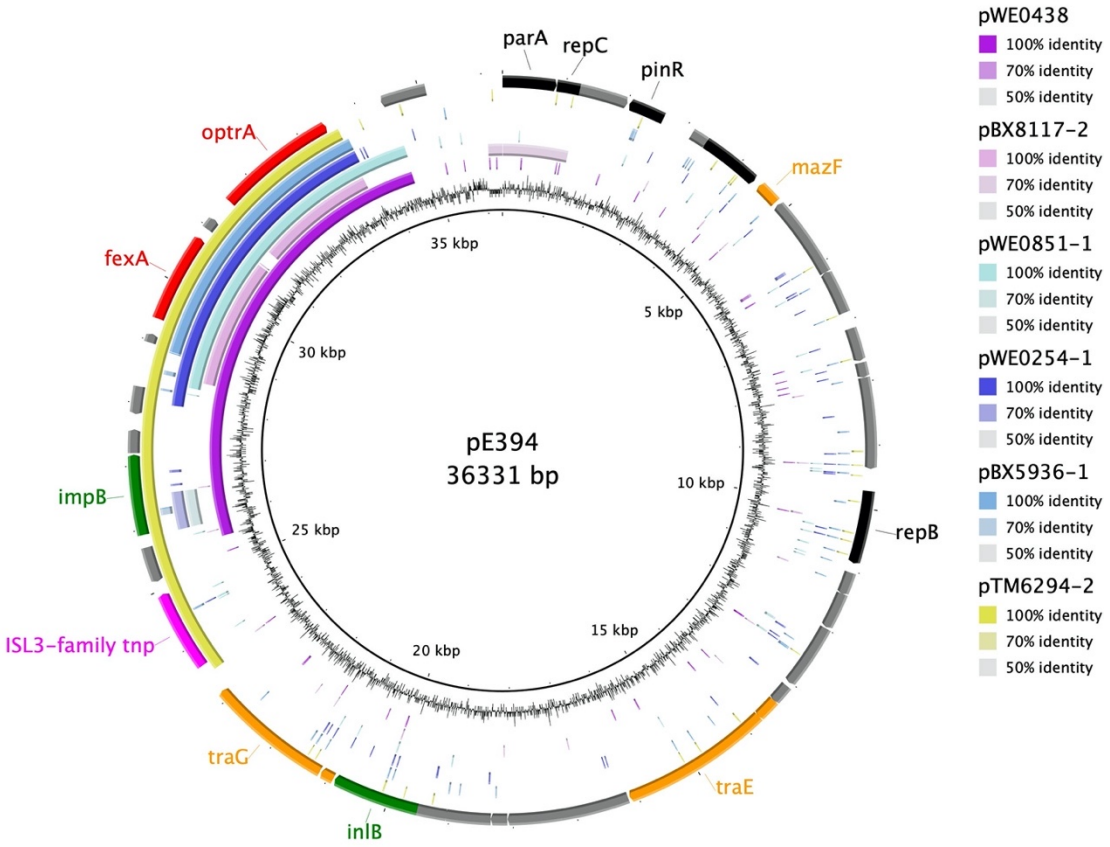


Figure 3.1 Alignment of full *optrA*-positive plasmid sequences against the first identified *optrA*-positive plasmid pE394

Sequence similarity confined to the *optrA*/*fexA* region. Inner ring indicates GC content of pE394, then alignment of pWE0438, pBX8117-2, pWE0851-1, pWE0254-1, pBX5936-1, pTM6294-2, and outer ring indicating CDSs in pE394 (accession KP399637) coloured by function: antimicrobial resistance (red), plasmid replication (black), transposition (pink), conjugative transfer (orange), plasmid maintenance or modification (green), and hypothetical protein (grey).

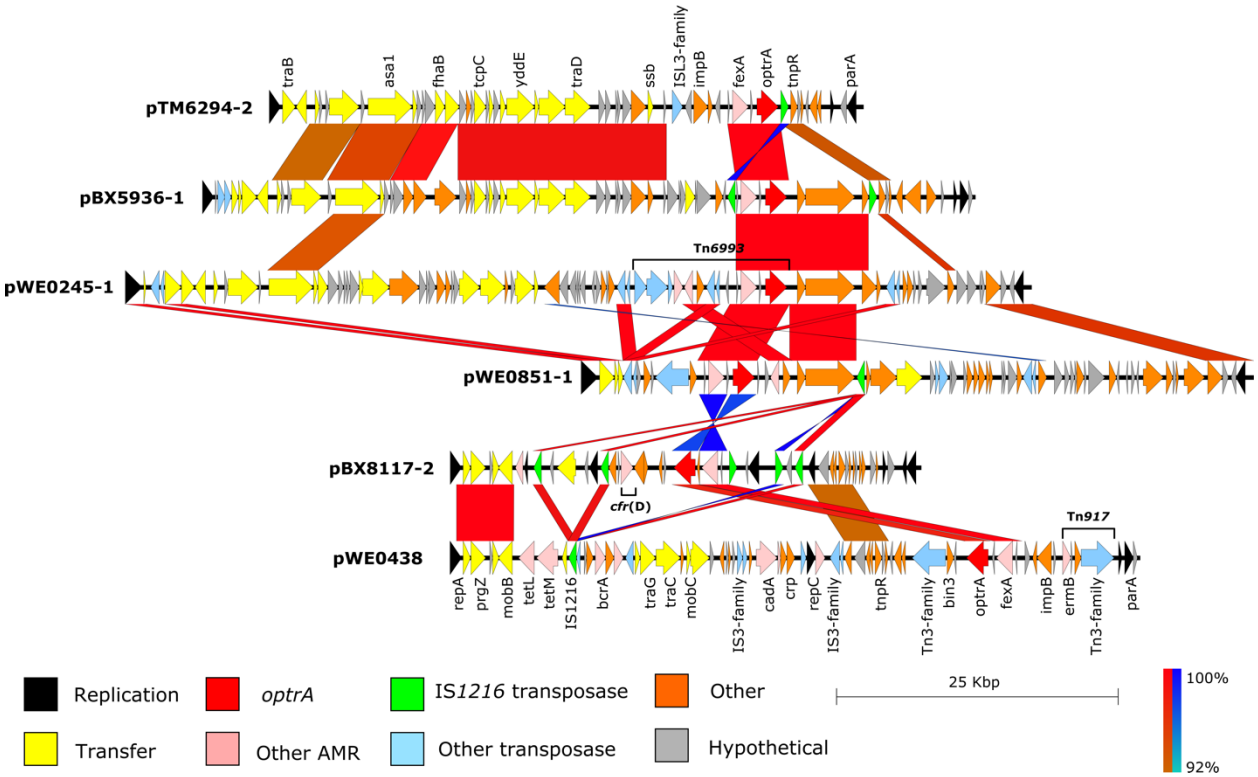


Figure 3.2 Alignment of full *optrA*-positive plasmid sequences to each other

While some sequence similarity is seen between pTM6294-2 and pBX5936-1, in general identity is low between the *optrA*-positive plasmids, indicating *optrA* has mobilised to multiple plasmid backbones. Arrows indicate CDSs, coloured blocks between each sequence indicate regions with BLASTn sequence identity ≥90% and length >680 bp. Blue identity blocks indicated inverted sequence.

IS*1216* is often associated with *optrA* and other AMR genes in enterococci. pBX5936-1 and pBX8117-2 had IS*1216* flanking the *optrA* and *fexA* region as a putative transposable cassette (Figure 3.2 and Figure 3.3). However, IS*1216* can mobilise from a single inserted copy [333] and single copies were found close to *optrA* in pTM6294-2 and pWE0851-1 (Figure 3.2 and Figure 3.3). BLASTn comparison of pWE0254-1 with the other *optrA*-positive plasmids highlighted a partial IS*1216* transposase that was not identified by automated annotation. Immediately upstream of the partial IS*1216* was an IS*3*-family transposase, the insertion of which likely disrupted IS*1216*. In pWE0254-1 *optrA* and *fexA* were found on a Tn*6674*-like element carrying macrolide (*ermA*) and spectinomycin (*aph*(9)-Ia) resistance genes. The element was 98.9% identical to Tn*6674* but had a 1.2 kb insertion containing IS*3*-family transposases (Figure 3.3), and was classified as Tn*6993* by the Transposon Registry (accession GCA_906464915) [334]. Tn*6993* was not inserted into the chromosomal *radC* gene as described for most Tn*6674*-like elements [335,336]. A similar element was present in a plasmid from *E. faecalis* in Chinese swine (TBCP-4814-p1, accession MH830363) but this element lacked the *tnpA* gene and the 1.2 kb insertion of Tn*6993* (Figure 3.3) [337]. pWE0438 had a single copy of IS*1216* located ~35 kb from *optrA*, although Tn*917* and Tn*3*-like transposases were detected closer to *optrA* as described above.

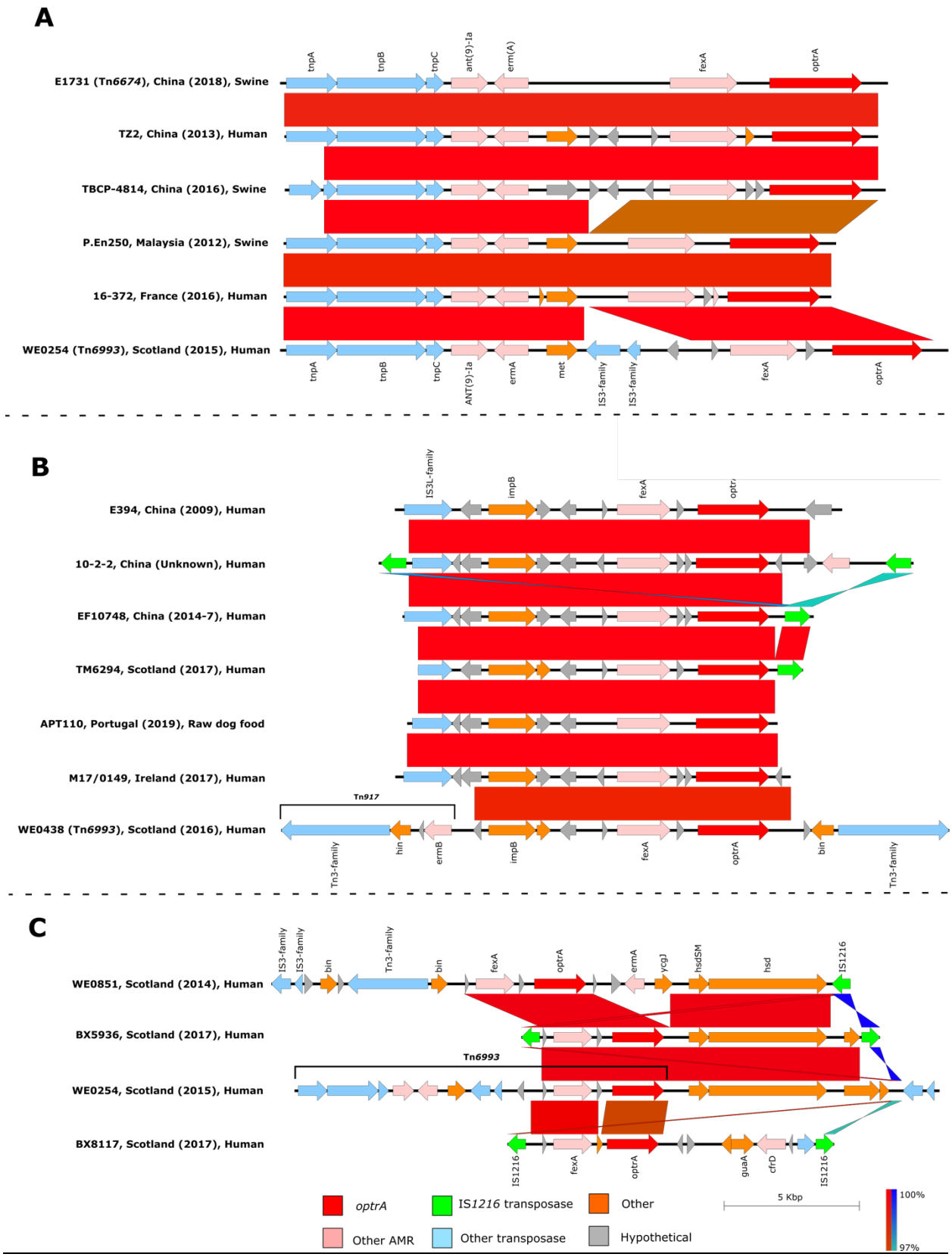

A
E1731 (Tn6674), China (2018), Swine
TZ2, China (2013), Human
TBCP-4814, China (2016), Swine
P.En250, Malaysia (2012), Swine
16-372, France (2016), Human
WE0254 (Tn6993), Scotland (2015), Human
tnpA
tnpB
tnpC
ant(9)-Ia
erm(A)
fexA
optrA
ermA
met
IS3-family
B
E394, China (2009), Human
10-2-2, China (Unknown), Human
EF10748, China (2014-7), Human
TM6294, Scotland (2017), Human
APT110, Portugal (2019), Raw dog food
M17/0149, Ireland (2017), Human
WE0438 (Tn6993), Scotland (2016), Human
IS3L-family
impB
Tn917
hin
ermB
Tn3-family
bin
C
WE0851, Scotland (2014), Human
BX5936, Scotland (2017), Human
WE0254, Scotland (2015), Human
BX8117, Scotland (2017), Human
IS3-family
ycgJ
hsdSM
hsd
IS1216
Tn6993
guaA
cfrD
optrA
IS1216 transposase
Other
Other AMR
Other transposase
Hypothetical
5 Kbp
100%
97%

Figure 3.3 Examples of different platforms carrying the *optrA* gene from diverse sources

Panel A includes examples of Tn*6674*-like platforms or Group 1 according to Freitas *et al* (2020); panel B includes *impB-fexA-optrA* platforms or Group 2; panel C includes the three isolates from this study that do not fit in the Freitas groupings, as well as WE0254 for comparison as it has conserved *hsd/hsdSM* restriction enzyme/methylase genes with WE0851 and BX5936 which may have a role in element maintenance. Arrows indicate CDSs, coloured blocks between each sequence indicate regions with >97% BLASTn sequence identity length >300 bp. Labels indicate the ID, country, year, and source of isolate.

3.2.3 *optrA* sequences vary between isolates

Comparison of the *optrA* sequence from each isolate to the first identified *optrA* from pE394 revealed different variants at the nucleotide and amino acid level: WE0254 and TM6294 had one synonymous nucleotide substitution, BX5936 had a single non-synonymous nucleotide substitution, WE0851 had two non-synonymous nucleotide substitutions, WE0348 had three non-synonymous and one synonymous substitution, and BX8117 had 20 non-synonymous and a further 17 synonymous substitutions (Appendices Table 3.1, Figure 3.4). In all cases, *optrA* and *fexA* were located within 550-750 bp, the degree of amino acid sequence variation identified in OptrA was not reflected in FexA sequences. Comparison to the first reported FexA sequence (AJ549214) showed four common non-synonymous variants in all strains (amino acid changes A34S, L39S, I131V, and V305I), with all but BX8117 having an additional D50A variant. This suggests that while there is evidence of diversification within *optrA* sequences, *fexA* is well conserved even when these two genes appear to be closely linked in the analysed genomes.

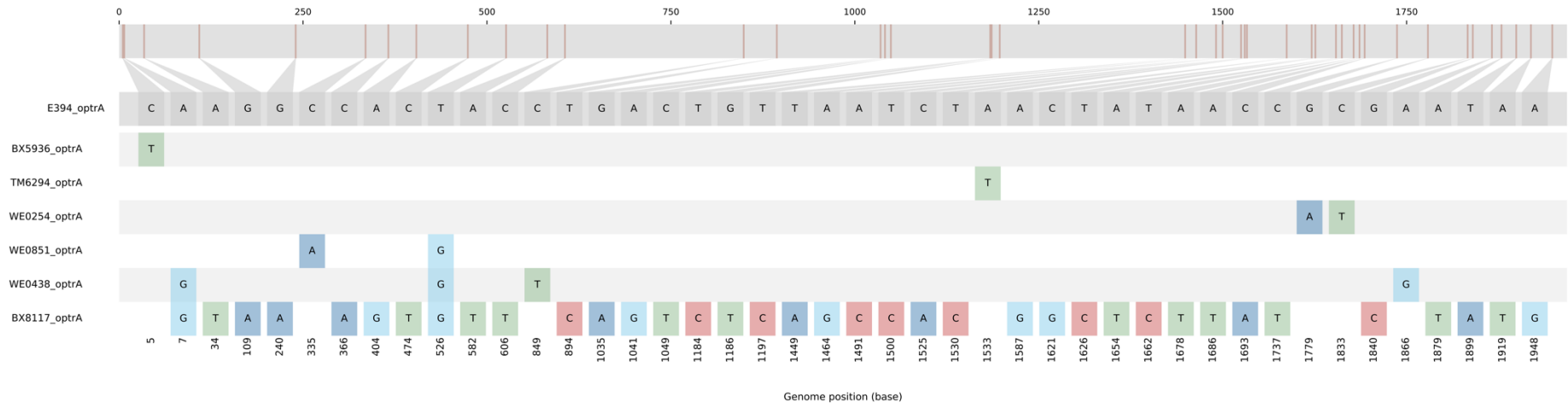


Figure 3.4 Nucleotide variants detected in Scottish *optrA* sequences

Variants against the first identified *optrA* identified in an *E. faecalis* isolated from a clinical sample in China in 2009 (pE394, accession KP399637).

3.2.4 *optrA*-positive strains are distantly related to UK bloodstream isolates

To investigate whether the *optrA*-positive isolates represented common *E. faecalis* strains in the UK, publicly available sequence data of 94 *E. faecalis* isolates from the British Society for Antimicrobial Chemotherapy (BSAC) bacteraemia surveillance programme (isolated between 2001 and 2011) were analysed together with the six known *optrA*-positive isolates.[250] We first looked for determinants of linezolid resistance in the 94 sequences, and found no evidence of *cfr*, *cfr*(B), *cfr*(D), *optrA*, *poxtA,* or the G2505A 23S rRNA gene mutation. Only one of the BSAC isolates carried the G2576T 23S rRNA gene mutation conferring linezolid resistance. Core genome phylogeny showed BX8117 was related to three other ST16 isolates from the UK, after removal of putative recombination blocks there were 76, 81, and 182 SNPs between these isolates suggesting they diverged from a common background but are not linked to recent transmission (Figure 3.5).

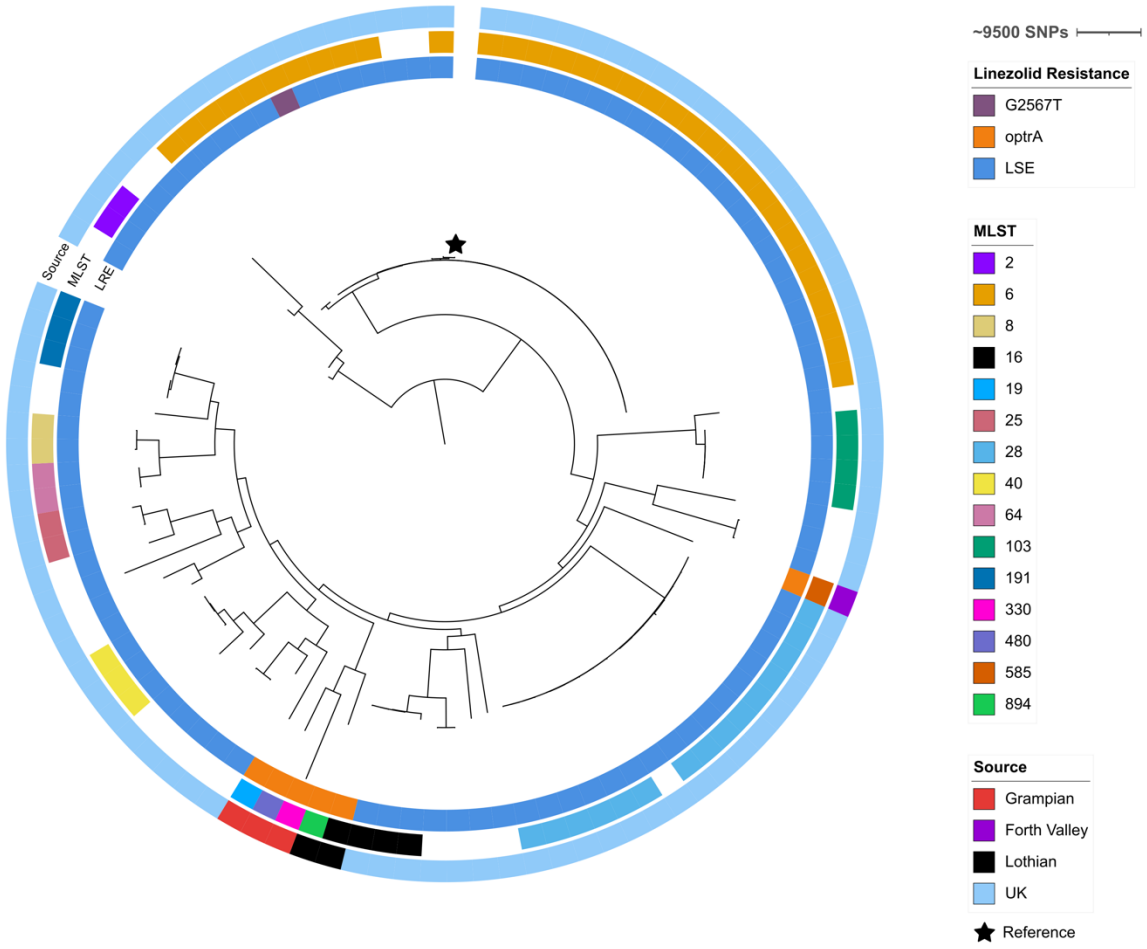


Figure 3.5 *optrA*-positive *E. faecalis* isolates in a national perspective

Phylogenetic analysis of the six *optrA*-positive isolates and 94 isolates from BSIs in the UK shows the *optrA*-positive isolates are generally unrelated to others in the collection. Illumina reads were mapped to *E. faecalis* V583 reference genome (3,218,031 bp), MGEs masked (701,199 bp), and a maximum likelihood phylogeny performed on SNP alignment (95,551 bp). Reference genome is highlighted by a star. Presence of linezolid resistance markers is indicated by the inner ring and coloured by resistance mechanism, the middle ring shows MLST for the Scottish *optrA*-positive isolates and any STs with two or more

cases in the BSAC collection, and the outer ring indicates isolate source. See figure for colour key.

3.3 Discussion

Until 2016, linezolid resistance was rare in Scottish enterococci (<1%), and when encountered was usually due to mutation of the 23S rRNA [81]. In 2016, an *optrA*-positive isolate was identified, and a lookback exercise identified further isolates in 2014 and 2015. The isolates included in this study represent some of the earliest identified *optrA*-positive enterococci from Scotland, investigation of which is important to identify potential means of introduction and transmission of this significant AMR mechanism. This study found *optrA* present in diverse genetic lineages of *E. faecalis* and carried on largely unrelated plasmids in six isolates from Scotland. pTM6294-2, pBX5936-1, and pWE0438 shared homology with plasmids identified in China and Japan, highlighting the wide dispersal of *optrA.* However, the other Scottish plasmids had limited similarity to other published examples suggesting a diverse reservoir of *optrA*-carrying genetic elements. In this study, *optrA* was often carried with several other resistance genes including in a novel multiresistance transposon Tn*6993* in pWE0254-1, and the recently described *cfr*(D) in pBX8117-2. Despite differences in *optrA* sequences and carriage of other linezolid determinants such as *cfr*(D), all isolates showed low level linezolid resistance of 8 mg/l (Table 3.1).

Comparison of the Scottish *optrA*-positive genomes with a UK-wide bacteraemia collection identified BX8117 clustered with ST16 isolates causing invasive disease (Figure 3.5). ST16 has been associated with multidrug-resistant infections in humans and animals,

highlighting the potential for the emergence of linezolid resistance in invasive enterococcal infections.[338] The other five *optrA*-positive isolates have no close genetic links in this phylogeny (minimum pairwise SNPs 12,314 – 17,891) suggesting they are generally distinct from those recently causing BSIs in the UK, however firm conclusions cannot be drawn as this study was not designed to infer national patterns.

Freitas *et al* [336] recently analysed all publicly available *optrA*-positive genome sequences and categorised the genetic environment of *optrA*. Group 1 includes Tn*6674*-like platforms, of which WE0254 is a representative (Figure 3.3). However, in the original scheme all Group 1 elements were integrated into the chromosome, in WE0254 the *optrA* element Tn*6993* is inserted into a plasmid. Group 2 includes *optrA*-*fexA*-*impB* platforms, represented in the Scottish isolates by TM6294 and WE0438 (Figure 3.3). Group 3 includes platforms containing the *araC* regulatory element and is not represented in the Scottish *optrA*-positive isolates characterised here. The three remaining Scottish isolates could not be grouped based on the Freitas scheme, highlighting the need for further studies and public access to complete genome sequences to determine the true diversity of *optrA*-positive platforms.

Many studies of *optrA* to date show a higher prevalence in animals (particularly in agriculture) compared to humans [78,339,340]. Further investigations have also identified *optrA*-positive isolates in samples of raw food in China, Colombia, Denmark, Poland, Switzerland, and Tunisia [341–347]. There are also reports of *optrA*-positive organisms in raw

food for companion animals in China, Portugal, and Switzerland, posing a risk for transmission from pets to humans [344,348,349]. Nüesch-Inderbinen *et al* [350] report 2.3% of healthy food processing plant workers in Switzerland carried *optrA*-positive enterococci and Cai *et al* [351] found 3.5% of healthy individuals in China carried *optrA*-positive enterococci. The European Food Safety Authority has identified *optrA* among the highest priority AMR mechanisms emerging and spreading through the food chain [352], highlighting the importance of hygiene practices, biosecurity, and food safety management. *optrA* confers resistance to linezolid and phenicols, and all the isolates described in this chapter also had the phenicol resistance gene *fexA* near *optrA*, phenicol use may provide a co-selection mechanism for *optrA*-mediated linezolid resistance. Indeed, the use of the phenicol florfenicol in agriculture is linked to *optrA* detection in farm animals [353,354]. Increasing reports describe *optrA* detection from human samples in many countries, although phenicol antibiotics are not widely used in human medicine [80,355,356]. *optrA*-positive isolates are often resistant to multiple antibiotic classes used in animal and human health, allowing significant opportunity for co-selection of *optrA*-positive strains in multiple settings. More recently, *optrA* has been identified in clinical vancomycin-resistant *E. faecium* isolates, with very limited treatment options [355,357,358].

This study is limited in scale as it only includes isolates from three regional clinical laboratories, therefore larger studies are required to infer national patterns. However, the finding that *optrA* is present as different gene variants, carried on different MGEs, in unrelated strains of *E. faecalis* suggest a diverse *optrA* reservoir that is only partly

investigated in this study. There is growing evidence that *optrA*-positive strains may be present in the food chain, and the pattern identified in Scottish *optrA*-positive genomes in this Chapter may reflect multiple introductions from the global food network.

As well as *optrA*, the *cfr* and *poxtA* genes are emerging transferable linezolid resistance mechanisms. Further studies from a One Health perspective are warranted to understand the selection pressures driving transferable linezolid resistance, and the transmission dynamics of these strains to avoid further spread of oxazolidinone resistance within *E. faecalis* and other Gram-positive bacteria.

Chapter 4 Consideration of within-patient diversity highlights transmission pathways and antimicrobial resistance gene variability in vancomycin resistant *Enterococcus faecium*

4.1 Introduction

Having investigated a cluster of isolates with a novel AMR mechanism, investigation of person-to-person transmission was performed next. As linezolid resistance is still rare in enterococci, focus switched to vancomycin resistance in *E. faecium* which is mainly a problem in hospital settings. In healthcare institutions asymptomatic intestinal carriage of VREfm can lead to shedding into the environment and transfer to other patients or staff, challenging efforts to limit the incidence of nosocomial infections.[359] WGS is increasingly applied to investigate transmission networks and identify control measures.[150,254] Many WGS based analyses of bacterial outbreaks, however, rely on analysing single colony picks from clinical samples assuming that this represents the entire infecting or colonizing population within individual patients.[360] It is increasingly recognised that within-patient diversity of bacterial populations can be significant and can influence transmission network resolution.[361–367] Several studies have identified that individual patients can carry multiple strains of *E. faecium* concurrently, but few have applied this to transmission resolution.[12,282,368–370]

The aim of this study was to identify how diverse the VREfm population is in rectal carriage and determine the optimal number of colonies to use to effectively detect transmission. A sampling strategy was designed to reliably detect within-patient diversity and supplemented short-read and long-read sequencing to generate high-quality reference genomes to identify genomic variants in the isolate collection.

Results presented in this chapter have been posted on the preprint server *medRxiv*: https://doi.org/10.1101/2022.09.23.22279632

4.2 Results

4.2.1 Epidemiological context

This study was performed over one month in 2017 on an inpatient unit for haematological malignancies, split into two wards (A and B). VREfm rectal screening was performed on all new admissions and any inpatients with febrile episodes to inform patient placement and antimicrobial administration. There was significant overlap between patient stays with some patients moving between the two study wards or to other wards in the hospital (Figure 4.1). Patients were cohorted or placed in single rooms when colonised with VREfm or other alert pathogens. However, not all rooms had ensuite bathroom facilities so risk of VREfm transmission remained. At the time of the study, surveillance systems in the hospital had not detected any suspected VREfm outbreak within the study population.

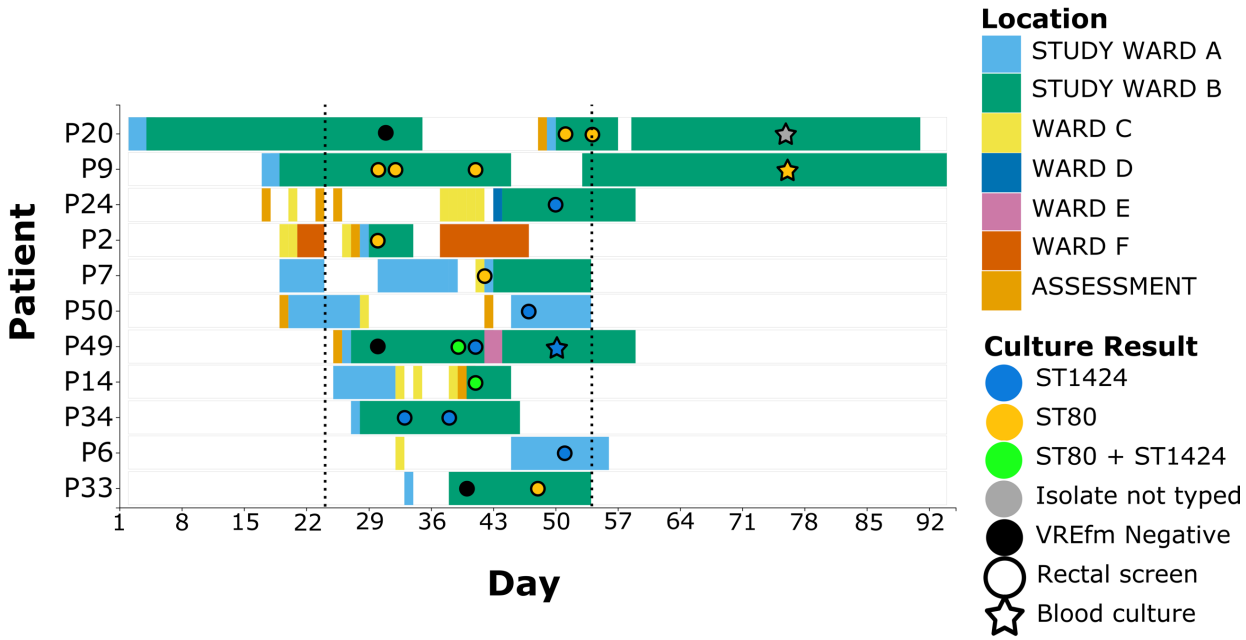


Figure 4.1 Patient timeline showing timing of ward stays and sample collection

Each row denotes the location of a patient during admission, blocks denote hospital stay, circles denote VREfm cultures, stars denote bloodstream isolates, dotted lines indicate the start and end of prospective collection of screening isolates for this study. This study was undertaken within Wards A and B, although patients were moved to different wards within the hospital during their stay and were often admitted through the assessment unit.

4.2.2 Design of sampling strategy

A sampling strategy was designed to reliably identify within-patient diversity from VREfm positive samples. The main considerations for detecting within patient diversity based on agar culture are the expected proportion of the population represented by a variant, the confidence required in the estimated prevalence, and the minimum number of colonies required to detect the given variant proportion with the given confidence. Power calculation was performed to estimate the minimum number of colonies required to identify a variant within the population at different proportions, with a confidence level of 95% (Table 4.1).

Table 4.1 Colonies required to identify population variant proportions with 95% confidence

Variant proportion in sample (%)	**Minimum colonies per sample required**
100	1
90	2
80	2
70	3
60	4
50	5
40	6

30	9
20	14
10	29
5	59
1	299

Next, the expected variant proportion in VREfm samples was estimated based on results published by Moradigaravand *et al* [282], a study of within-patient diversity in four patients undergoing longitudinal stool carriage surveillance who developed BSI. Publicly available reads of VREfm from the study were mapped to the Aus0004 reference genome and pairwise SNPs counted to identify population variants within samples. Multiple colony picks were analysed from 11 stool samples which found clonal populations in three samples and 2-3 variants in the other eight samples with the minor variant accounting for 20-50% of the population (Table 4.2). Variants could be easily identified as they differed from each other by a median 4964 (range 3798-9773) SNPs and were generally different MLST STs, diversity was much lower within population variants (median 2 SNPs, range 0-6 SNPs). Moradigaravand *et al* [282] used five or eight colony picks in these samples which would be expected to reliably detect variants accounting for 50% or 32% of the population, respectively (Table 4.1). BSI populations were also analysed by sequencing a median of 10 (range 2-18) colonies from eight blood cultures, but even with the higher

power this afforded population diversity was low with median pairwise difference of 1 (range 0-15) SNP and no population variants identified.

Table 4.2 Population diversity previously identified in rectal samples[a]

Sample ID	Colonies analysed	Variants detected	Minor variant proportion of total (%)
B-0	5	1	100
C-85	5	1	100
D-172	5	1	100
B-9	8	2	50
A-130	5	2	40
B-14	5	2	40
B-52	5	2	40
C-99	5	2	40
C-134	5	2	20
A-138	5	3	20
B-20	5	3	20

[a] Reference Moradigaravand *et al* [282]

Based on the findings described above, it was deemed necessary to accurately identify a minor variant comprising 20% of the rectal population as this was the lowest proportion identified by Moradigaravand *et al* [282]. The minimum number of colonies to analyse was

therefore identified as 14 per rectal sample. For positive blood cultures, the available evidence suggests these are very clonal populations (either due to a single clonal seeding event of the bloodstream, or a bias introduced during blood culture) and so only one colony was analysed in this study.

4.2.3 Results of VREfm screening

In total, 45 rectal swabs from 27 patients were screened for VREfm. Of these, 18 samples from 13 patients were VREfm positive (Table 4.3). Three (23.1%) colonised patients developed VREfm bacteraemia 9, 24, or 46 days after being identified as VREfm carriers. Two rectal swabs and one blood culture were not available for further study. The sampling strategy of picking 14 random colonies was applied to 16 rectal screens from 11 patients generating 224 isolates. Bloodstream isolates were available from five blood cultures from two patients, the addition of single colonies from these five cultures produced a total of 229 isolates. Most patients were female, the median age was 66 years, and a range of primary diagnoses were present (Table 4.3). Most colonised patients had received antibiotics in the preceding six months and 30% had received vancomycin (Table 4.3).

Table 4.3 Characteristics of patients with rectal VREfm colonisation (n = 13)

Demographics	**Number (%)**
Female	8 (61.5)

Age, median (range) years	66 (37-77)
Primary diagnosis	
Acute myeloid leukaemia	3 (23.1)
Diffuse large B cell lymphoma	3 (23.1)
Multiple myeloma	3 (23.1)
Myelodysplasia	2 (15.4)
Chronic lymphocytic leukaemia	1 (7.7)
Mantle cell lymphoma	1 (7.7)
Antimicrobial administration	
Any antibiotics in the 7 days prior to positive screen	12 (92.3)
Any antibiotics in the 6 months prior to positive screen	12 (92.3)
Vancomycin in the 7 days prior to positive screen	1 (7.7)
Vancomycin in the 6 months prior to positive screen[a]	3 (30.0)
Outcomes within 60 days of VREfm positive screen	
VREfm BSI	3 (23.1)
Intensive care unit admission	1 (7.7)
Death	0 (0)

[a] Information available for 10 patients

BSI, bloodstream infection

4.2.4 Simultaneous carriage of multiple VREfm strains

In silico MLST typing using short reads from all 229 genomes showed ST80 (n=130), ST1424 (n=97), ST789 (n=1), and ST1659 (n=1) from the hospital-associated clade A[371] were present (Table 4.4). Multiple STs were detected in three (27%) samples. Sample VRED06 from patient P49 contained 10 (71.4%) ST80, three (21.4%) ST1424, and one (7.1%) ST789 isolate; sample VRED07 from P14 contained 10 (71.4%) ST1424 and four (28.6%) ST80 isolates; sample VRED11 from P50 contained 13 (92.9%) ST1424 and one (7.1%) ST1659 isolate. A further rectal swab sample from P49 collected two days after VRED06 contained only ST1424, and a blood culture collected nine days later also contained ST1424. P9 had three rectal swab samples collected over 11 days and had positive blood cultures one month later, all samples contained ST80 only. Within-patient populations could be identified based on assigned ST, within STs genetic distance was low with a maximum pairwise distance of 3 SNPs (median 0 SNPs, range 0-3 SNPs; Table 4.4).

Table 4.4 STs detected within patients

Patient ID	Sample ID	Sample Type	Sample Date (days from start of study)	STs detected (n, %)	Maximum pairwise SNP distance within sample	Median (IQR) pairwise SNP distance within sample
P2	VRED01	Rectal	6	80 (14, 100)	2	0 (0 - 1)
P6	VRED16	Rectal	27	1424 (14, 100)	2	0 (0 - 1)
P7	VRED10	Rectal	18	80 (14, 100)	2	0 (0 - 1)
P9	VRED02	Rectal	6	80 (14, 100)	2	0 (0 - 0)
P9	VRED03	Rectal	8	80 (14, 100)	0	0 (0 - 0)
P9	VRED09	Rectal	17	80 (14, 100)	0	0 (0 - 0)
P9	VRED18	Blood	52	80 (1, 100)	-	-
P9	VRED19	Blood	52	80 (1, 100)	-	-
P9	VRED20	Blood	52	80 (1, 100)	-	-
P9	VRED21	Blood	52	80 (1, 100)	-	-
P14	VRED07	Rectal	17	1424 (10, 71.4) 80 (4, 26.6)	2 2	1 (0-1) 2 (1-2)
P20	VRED15	Rectal	27	80 (14, 100)	3	0 (0 - 2)
P20	VRED17	Rectal	30	80 (14, 100)	3	1 (0 - 1)
P24	VRED13	Rectal	26	1424 (14, 100)	0	0 (0 - 0)
P33	VRED12	Rectal	24	80 (14, 100)	1	0 (0 - 0)
P34	VRED04	Rectal	9	1424 (14, 100)	0	0 (0 - 0)
P34	VRED05	Rectal	14	1424 (14, 100)	1	0 (0 - 0)

P49	VRED06	Rectal	15	80 (10, 71.4) 1424 (3, 21.4) 789 (1, 7.1)	2 1 -	0 (0 - 1) 0 (0 - 0) -
P49	VRED08	Rectal	17	1424 (14, 100)	2	0 (0 - 1)
P49	VRED14	Blood	30	1424 (1, 100)	-	-
P50	VRED11	Rectal	23	1424 (13, 92.9) 1659 (1, 7.1)	3 -	1 (0 - 2) -

ID, identification; IQR, interquartile range

4.2.5 Genomic population structure of VREfm suggests recent transmission events

The chromosomes of the two strain-specific genome assemblies (VRED06-02 and VRED06-10, Table 4.5, and Appendix 2) were used as references for short-read mapping within each ST. Within-patient diversity was low when genomes of the same ST were compared, generally differing by zero SNPs and a maximum pairwise difference of 3 SNPs (Table 4.4). Similarly, insertions, deletions, and plasmids were usually shared in genomes from the same patient. However, the presence of DEL3 (12 bp non-coding deletion) and DEL4 (11 bp deletion in *adcA* encoding solute binding protein accession WP_002297324) were variable within 24 ST80 genomes from P20 with 0-2 differentiating SNPs (Figure 4.2). In genomes from P9 p1_VRED06-10 and p3_VRED06-10 were variably detected despite most genomes having no differentiating SNPs (Figure 4.2).

Table 4.5 Quality metrics for genome assemblies

Isolate	**MLST**	**BUSCO [n, (%)]**				**Truncated CDS, possible indel errors [n, (%)]**	
		Total BUSCOS	**Complete**	**Fragmented**	**Missing**	**Total CDS**	**CDS <0.9x reference**
VRED06-02	1424	402	400 (99.5)	1 (0.2%)	1 (0.2)	3163	122 (3.4)
VRED06-10	80	402	400 (99.5)	1 (0.2%)	1 (0.2)	3122	124 (3.9

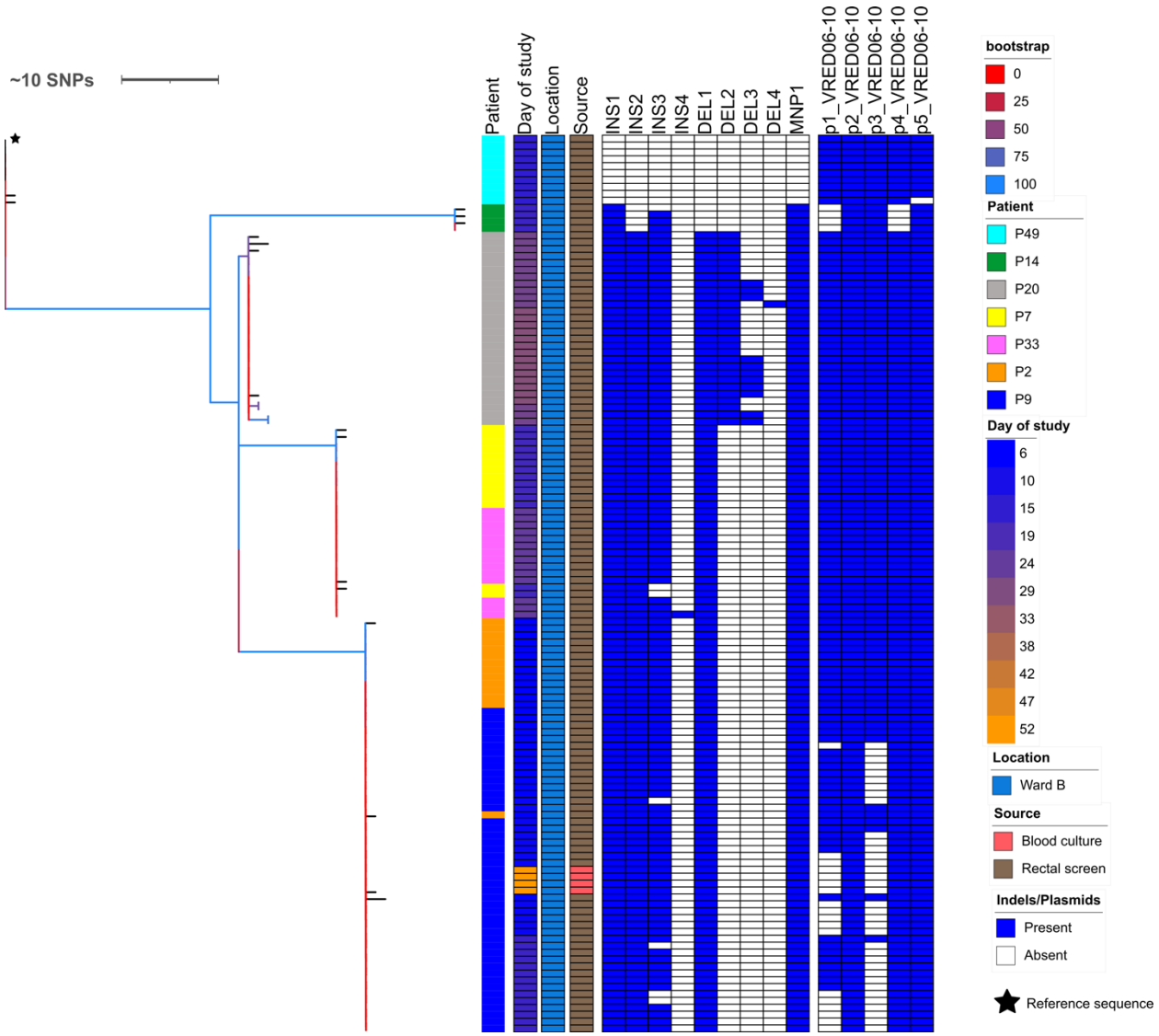


Figure 4.2 Phylogeny of ST80 within-patient isolates

Shows structured population with three patient specific clusters and two clusters indicating putative patient-to-patient transmission of VREfm. All ST80 isolates (n=130) mapped to VRED06-10 chromosome (2,814,943 bp), MGEs and recombination masked (202,738 bp) and maximum likelihood phylogeny built on remaining SNPs (96 bp).

Metadata is indicated by coloured blocks (see figure for key). Tree unrooted. DEL, deletion; INS, insertion; MNP, multiple nucleotide polymorphism

The ST80 genomes formed a well-structured population with five clear clusters each separated by >10 SNPs (Figure 4.2). Clustered genomes differed by 0-2 SNPs and were mostly from individual patients although two clusters included genomes from two different patients (patients P7 and P33, and P2 and P9). All the reference plasmids were detected in the P7 and P33 genomes, considering insertions INS1 and INS2 were present in all genomes while INS3 was present in all but two genomes from P7. INS4 was additionally only detected in a single genome from P33. Within P9 genomes, p1_VRED06-10 was detected in 28/48 and p3_VRED06-10 in 10/48, although in P2 genomes all plasmids were detected.

Mapping of the ST1424 genomes showed a much more homogeneous population than in ST80 (Figure 4.3). Of the 97 ST1424 genomes, 69 had no SNPs and the remaining 28 had 1-2 SNPs differentiating them from the rest of the collection. The SNPs that were detected did not lead to any clear clustering of genomes, except for the 14 genomes from P6 which all carried a SNP in a penicillin-binding protein which differentiated them from the other ST1424 genomes. Two of the P6 genomes had further independent SNPs (one each) and another genome had lost p1_VRED06-02. No insertions were detected in the ST1424 collection, and of the six deletions found five were only in genomes from P49. p6_VRED06-02 was not detected in 14 P24, 14 P50, and two P49 genomes, while p1_VRED06-02 was not detected in five genomes from three patients.

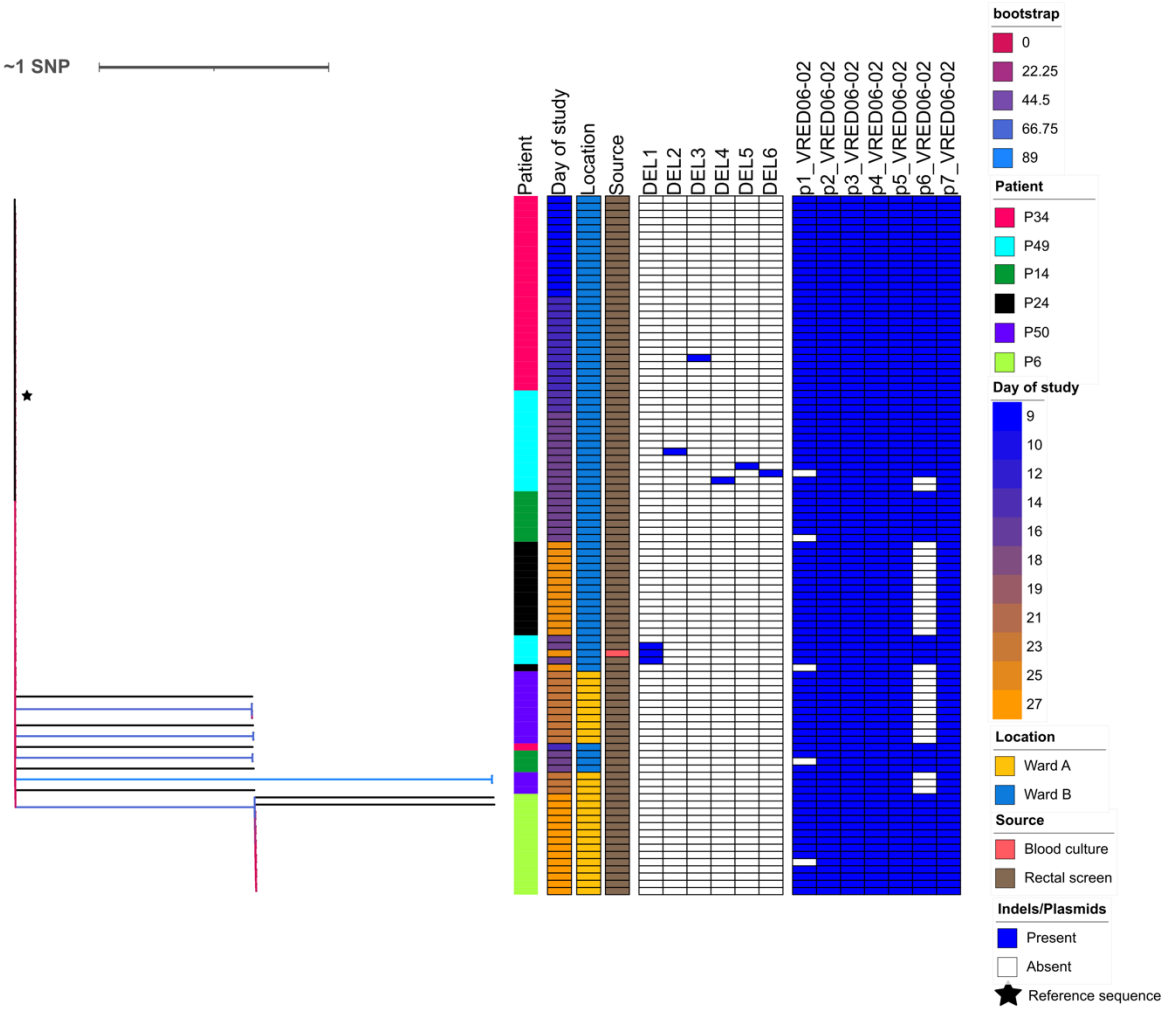


Figure 4.3 Phylogeny of within-patient ST1424 genomes

Shows homogeneous population suggestive of recent transmission outbreak. ST1424 genomes (n=97) mapped to VRED06-02 chromosome (2,945,113 bp), MGEs and recombination masked (227,540 bp) and maximum likelihood phylogeny built on remaining SNPs sites (13 sites). Metadata is indicated by coloured blocks (see figure for key). Tree unrooted. DEL, deletion

4.2.6 Analysis of multiple VREfm colonies supports transmission resolution

Transmission identification was performed with Phyloscanner after mapping all 229 genomes to the VRED06-10 ST80 reference. Phyloscanner takes multiple phylogenetic trees containing host origin information, determines ancestral host states, and then identifies putative transmission events between individual hosts. The final transmission network based on all 229 genomes is Figure 4.4. The network supports transmission of ST80 between P2 and P9, and between P7 and P33, with P20 not linked to transmission. Epidemiological data supports transmission from P7 to P33 on Ward B, as P33 screened negative early in their admission and then screened positive six days after P7 (Figure 4.1, Figure 4.4). P9 and P2 screened positive on the same day - no shared rooms or bed spaces were identified as this was P2's first day on Ward B so it is unclear where or when transmission may have occurred (Figure 4.1, Figure 4.5). P20 had two admissions during the study period, was negative at the end of first admission then screened positive on re-admission suggesting they may have become colonised outside of the hospital.

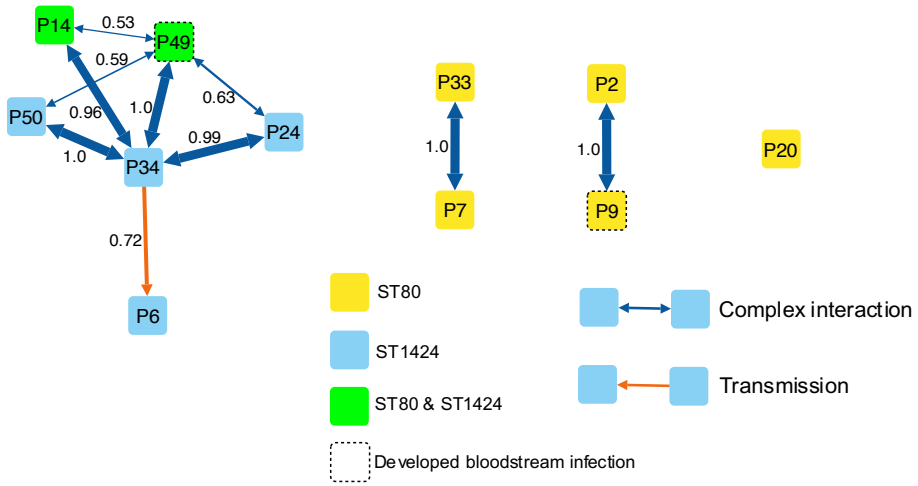


Figure 4.4 Transmission network showing putative links between patients

Each patient is represented by a node coloured by detection of the two outbreak STs. Edge thickness corresponds to fraction of Phyloscanner trees with given relationship, relationship fraction is printed alongside each edge, and edge colour based on type of relationship (orange, direct transmission; blue, transmission but direction unclear). Interactions were defined with Phyloscanner based on 100 random MrBayes trees and the network visualised in Cytoscape.

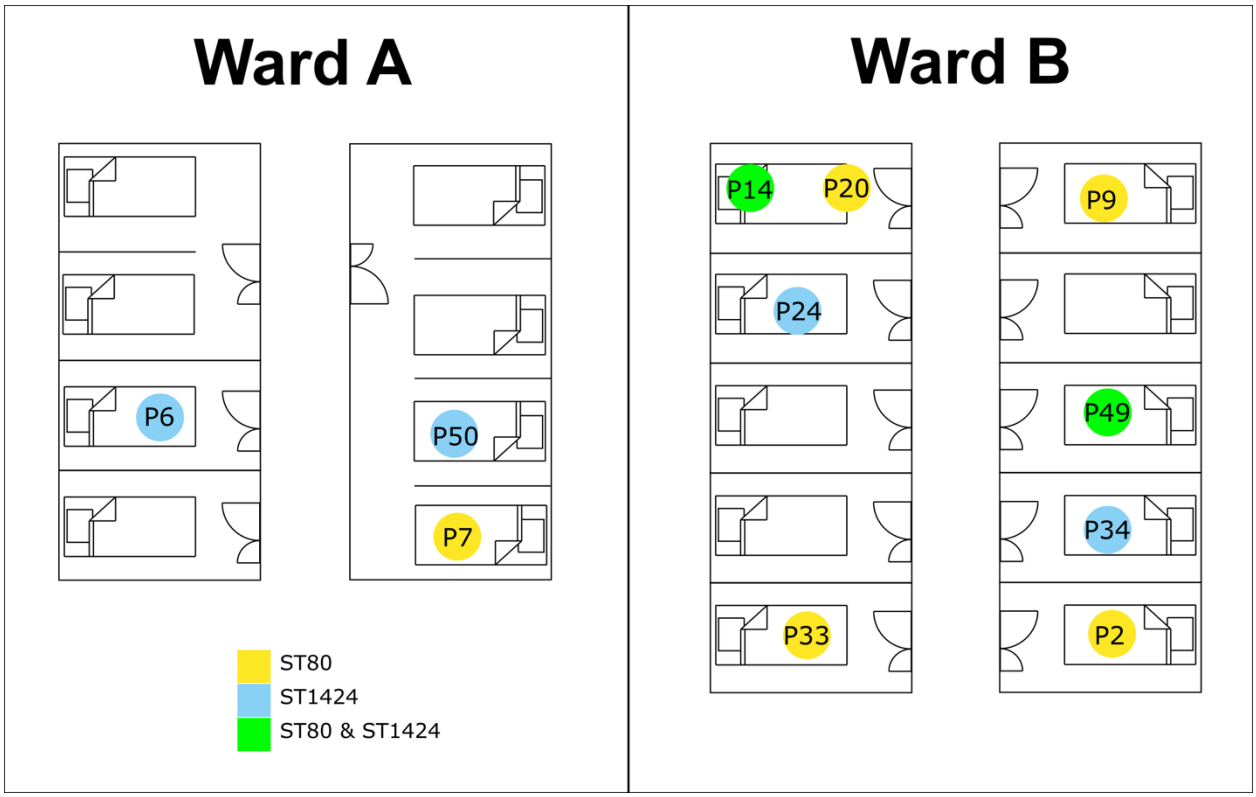


Figure 4.5 Patient locations at time of screening positive for VREfm

Schematic diagram of patient placement on Wards A and B at the time of first testing positive for VREfm during the study. Ward A comprised one four-bedded room, one two-bedded room, and two one-bedded rooms, Ward B comprised 10 one-bedded rooms. Patients are indicated by circles, coloured by the main STs identified within each patient.

All ST1424 patients clustered together with P34 strongly linked to all patients and likely direct transmission to P6 (Figure 4.4). P34 was the first ST1424 identified on Ward B, P49 was positive six days later (having been negative earlier in admission), P14 was positive two days after that, and P24 was positive 9 days subsequently (Figure 4.1). P14 and P49 had ST1424 and ST80 in carriage samples, sharing of the ST80 lineages in these patients was not identified suggesting there was no direct transmission between these two patients. On Ward A, P50 screened positive with ST1424 and ST1659 on day two of admission and P6 screened positive for ST1424 on day six. The ST1424 populations in P6 and P50 may derive from different hosts with P6 genomes all having a single SNP and P50 genomes having multiple different SNPs and lack the p6_VRED06-02 plasmid (Figures 4.3 and 4.5). P34 and P50 shared time on Ward A early in the study before either were known to be VREfm positive, but there is very limited overlap in time between P34 and P6 while both were in different wards (Figure 4.1). None of the patients with ST1424 shared a room or used a bed space previously used by an identified ST1424-positive carrier during their stay (Figure 4.5).

Analysing less than 14 colonies per sample produced fewer transmission links and lower confidence (Table 4.6, Figure 4.6). Linkage within the ST80 clusters was strong in all cases, but within ST1424 it was more variable. For example, with three colonies P14 was weakly associated with transmitting to P6 and P49, with five colonies P49 was weakly associated with P50 and P6 was unlinked, and with 10 colonies P49 was weakly associated with P24, moderately associated with P34, and P6 was unlinked. With 14 colonies, there was

moderate support for direct transmission from P34 to P6 and P34 was strongly associated with all other cases, which could be supported by the epidemiological data.

Table 4.6 Transmission network scores for different colony thresholds

	3 Colonies	**5 Colonies**	**10 Colonies**	**14 Colonies**
Total Colonies	30	50	100	229
Transmissions detected	9	9	10	10
Transmission confidence, median (min-max)	0.84 (0.58-0.99.0)	0.73 (0.57-1.0)	0.90 (0.53-1.0)	0.98 (0.53-1.0)

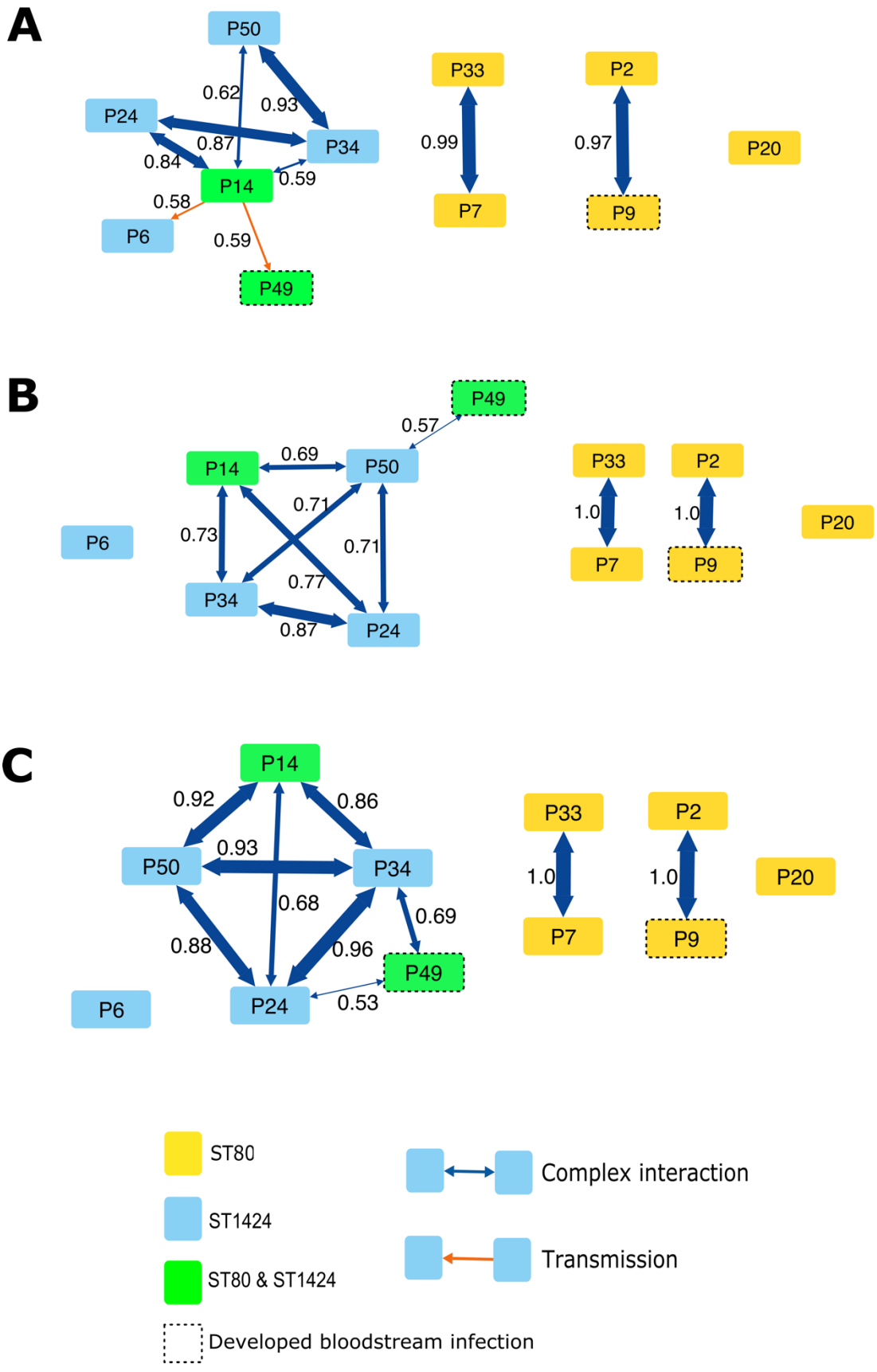

A
P50
0.62
0.93
P24
0.87
P34
0.84
0.59
P14
0.58
P6
0.59
P49
P33
0.99
P7
P2
0.97
P9
P20
B
P49
0.57
P14
0.69
P50
0.71
P6
0.73
0.71
0.77
P34
0.87
P24
P33
1.0
P7
P2
1.0
P9
P20
C
P14
0.92
0.86
0.93
P50
P34
0.68
0.69
0.88
0.96
P49
P6
P24
0.53
P33
1.0
P7
P2
1.0
P9
P20
ST80
ST1424
ST80 & ST1424
Developed bloodstream infection
Complex interaction
Transmission

Figure 4.6 Effect of different sampling strategies on transmission inference

Phyloscanner transmission networks for three (A), five (B), and ten (C) colony picks. Edge thickness corresponds to fraction of Phyloscanner trees with given relationship, relationship fraction is printed alongside each edge, and edge colour based on type of relationship (orange, direct transmission; blue, transmission but direction unclear). Relationships were determined with Phyloscanner based on 100 random MrBayes trees and the networks visualised with Cytoscape.

4.2.7 Plasmids were mostly ST-specific

Given the detection of distinct STs within P49 and the availability of high-quality plasmid assemblies, it was then investigated whether there was any evidence of within patient plasmid transfer. VRED06-02 (ST1424 reference) contained seven plasmids, and VRED06-10 (ST80 reference) contained five plasmids. Plasmid sequences were compared using Mash to identify similarity, with a score of 0 being identical and 1 being highly dissimilar. Plasmids in the two genomes were generally distinct, suggesting limited sharing between STs within P49 (Table 4.7).

Table 4.7 Mash distance of reference isolate plasmids

	p1_VRED06-02	p2_VRED06-02	p3_VRED06-02	p4_VRED06-02	p5_VRED06-02	p6_VRED06-02	p7_VRED06-02	p1_VRED06-10	p2_VRED06-10	p3_VRED06-10	p4_VRED06-10	p5_VRED06-10
p1_VRED06-02	0.00	0.15	0.13	1.00	1.00	1.00	1.00	0.06	0.20	0.11	1.00	1.00
p2_VRED06-02	0.15	0.00	0.17	0.30	0.16	0.30	0.16	0.16	0.09	0.09	0.23	0.30
p3_VRED06-02	0.13	0.17	0.00	1.00	1.00	1.00	1.00	0.11	0.18	0.09	1.00	1.00
p4_VRED06-02	1.00	0.30	1.00	0.00	0.18	1.00	1.00	1.00	1.00	0.24	0.05	0.16
p5_VRED06-02	1.00	0.16	1.00	0.18	0.00	0.15	1.00	1.00	1.00	0.24	0.15	0.11
p6_VRED06-02	1.00	0.30	1.00	1.00	0.15	0.00	0.17	1.00	1.00	0.26	1.00	1.00
p7_VRED06-02	1.00	0.16	1.00	1.00	1.00	0.17	0.00	1.00	1.00	0.24	1.00	1.00
p1_VRED06-10	0.06	0.16	0.11	1.00	1.00	1.00	1.00	0.00	1.00	0.17	1.00	1.00
p2_VRED06-10	0.20	0.09	0.18	1.00	1.00	1.00	1.00	1.00	0.00	1.00	1.00	1.00
p3_VRED06-10	0.11	0.09	0.09	0.24	0.24	0.26	0.24	0.17	1.00	0.00	0.23	0.26

p4_VRED06-10	1.00	0.23	1.00	0.05	0.15	1.00	1.00	1.00	1.00	0.23	0.00	0.18
p5_VRED06-10	1.00	0.30	1.00	0.16	0.11	1.00	1.00	1.00	1.00	0.26	0.18	0.00

Coloured based on similarity: ≤0.01, green; ≤0.05, yellow; ≤0.1, blue.

Carriage of similar plasmids was sought in the entire collection by short read mapping (Table 4.8). Most plasmids were ST-specific with few examples of ST1424 genomes carrying plasmids from the ST80 reference, and *vice versa*. However, all ST80 genomes from P7 and P33 carried p7_VRED06-02 from ST1424, and almost all genomes appeared to carry p4_VRED06-10. The hits against the ST1424 genomes are likely due to cross-mapping of reads from the related p4_VRED06-02 (Table 4.7). P7_VRED06-02 is unrelated to others in the collection (Table 4.7), but no close links to any ST1424-positive patients were identified for P7 and P33 (Figure 4.4).

Table 4.8 Plasmid detection in complete genome collection (n=229)

Patient	STs (n)	ST1424 Reference Plasmids							ST80 Reference Plasmids				
		p1_VRED06-02	p2_VRED06-02	p3_VRED06-02	p4_VRED06-02	p5_VRED06-02	p6_VRED06-02	p7_VRED06-02	p1_VRED06-10	p2_VRED06-10	p3_VRED06-10	p4_VRED06-10[a]	p5_VRED06-10
P34	1424 (28)	28 (100)	28 (100)	28 (100)	28 (100)	28 (100)	28 (100)	28 (100)	0 (0)	0 (0)	0 (0)	28 (100)	0 (0)
P6	1424 (14)	13 (92.9)	14 (100)	14 (100)	14 (100)	14 (100)	14 (100)	14 (100)	0 (0)	0 (0)	0 (0)	14 (100)	0 (0)
P24	1424 (14)	13 (92.9)	14 (100)	14 (100)	14 (100)	14 (100)	0 (0)	14 (100)	0 (0)	0 (0)	0 (0)	14 (100)	0 (0)
P50	1424 (13)	13 (100)	13 (100)	13 (100)	13 (100)	13 (100)	0 (0)	13 (100)	0 (0)	0 (0)	0 (0)	13 (100)	0 (0)
	1659 (1)	1 (100)	1 (100)	0 (0)	1 (100)	1 (100)	0 (0)	1 (100)	0 (0)	0 (0)	0 (0)	1 (100)	0 (0)

P14	1424 (10)	8 (80.0)	10 (100)	10 (100)	10 (100)	10 (100)	10 (100)	10 (100)	0 (0)	0 (0)	0 (0)	10 (100)	0 (0)
	80 (4)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	4 (100)	4 (100)	0 (0)	4 (100)
P49	1424 (18)	17 (94.4)	18 (100)	18 (100)	18 (100)	18 (100)	16 (88.9)	18 (90.9)	0 (0)	0 (0)	0 (0)	18 (100)	0 (0)
	80 (10)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	10 (100)	10 (100)	10 (100)	10 (100)	9 (90.9)
	789 (1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (100)	1 (100)	1 (100)	1 (100)	1 (100)
P7	80 (14)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	14 (100)	14 (100)	14 (100)	14 (100)	14 (100)	14 (100)
P20	80 (28)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	28 (100)	28 (100)	28 (100)	28 (100)	28 (100)

P2	80 (14)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	14 (100)	14 (100)	14 (100)	14 (100)	14 (100)
P33	80 (14)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	14 (100)	14 (100)	14 (100)	14 (100)	14 (100)	14 (100)
P9	80 (46)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	28 (60.9)	46 (100)	10 (21.7)	46 (100)	46 (100)

[a] p4_VRED06-10 is shorter than but homologous to p4_VRED06-02, the matches in P6, P24, and P34, and ST1424/1659 P14, P49, and P50 genomes are likely false positives.

4.2.8 AMR gene content differs between closely related genomes

The variability of AMR genes within the collection was investigated next (Table 4.9 and Figure 4.7). In total 13 AMR genes were detected with three (*aac(6')-Ii*, *msr(C)*, and *vanA*) present in all genomes, two (*aph(3')-III* and *erm(B)*) in all but one genome, four (*ant(9)-Ia*, *dfrG*, *erm(A)*, and *tet(M)*) only in ST1424 or ST1659 genomes, two genes (*ant(6)-Ia* and *tet(S)*) found only in ST80 and ST789 genomes, and *tet(L)* found in a single ST1659 genome. The aminoglycoside resistance gene *aac(6')-aph(2'')* was variably present, found in 69.9% of all genomes.

Table 4.9 Presence of AMR Genes in complete genome collection (n=229)

Gene	Phenotypic resistance	ST80, n = 130		ST1424, n = 97		All Genomes, n = 229	
		n (%)	Genetic element	n (%)	Genetic element	n (%)	Summary
aac(6')-aph(2'')	Amikacin, Gentamicin, Kanamycin, Streptomycin, Tobramycin	65 (50.0)	p1_VRED06-10	94 (96.9)	p1_VRED06-02	160 (69.9)	Variable in ST80/ST1424/ST789
aac(6')-Ii	Gentamicin, Tobramycin	130 (100)	Chromosome	97 (100)	Chromosome	229 (100)	All genomes
ant(6)-Ia	Streptomycin	130 (100)	p3_VRED06-10	0 (0)	-	131 (57.2)	All ST80/ST789
ant(9)-Ia	Spectinomycin	0 (0)	-	97 (100)	Chromosome	97 (42.4)	All ST1424
aph(3')-III	Amikacin, Kanamycin, Neomycin	130 (100)	p3_VRED06-10	97 (100)	p2_VRED06-02	228 (99.6)	All ST80/ST789/ST1424
dfrG	Trimethoprim	0 (0)	-	97 (100)	Chromosome	97 (42.4)	All ST1424
erm(A)	Clindamycin, Erythromycin, Quinupristin	0 (0)	-	97 (100)	Chromosome	97 (42.4)	All ST1424
erm(B)	Clindamycin, Erythromycin, Quinupristin	129 (99.2)	p1_VRED06-10, p3_VRED06-10	97 (100)	p2_VRED06-02	228 (99.6)	All except one ST80 genome
msr(C)	Erythromycin, Quinupristin	130 (100)	Chromosome	97 (100)	Chromosome	229 (100)	All genomes
tet(L)	Doxycycline, Tetracycline	0 (0)	-	0 (0)	-	1 (0.4)	Only ST1659
tet(M)	Doxycycline, Minocycline, Tetracycline	0 (0)	-	60 (61.9)	Chromosome	61 (26.6)	Variable in ST1424/ST1659
tet(S)	Doxycycline, Minocycline, Tetracycline	130 (100)	p3_VRED06-10	0 (0)	-	131 (57.2)	All ST80/ST789
vanA	Teicoplanin, Vancomycin	97 (100)	p2_VRED06-10	97 (100)	p2_VRED06-02	229 (100)	All genomes

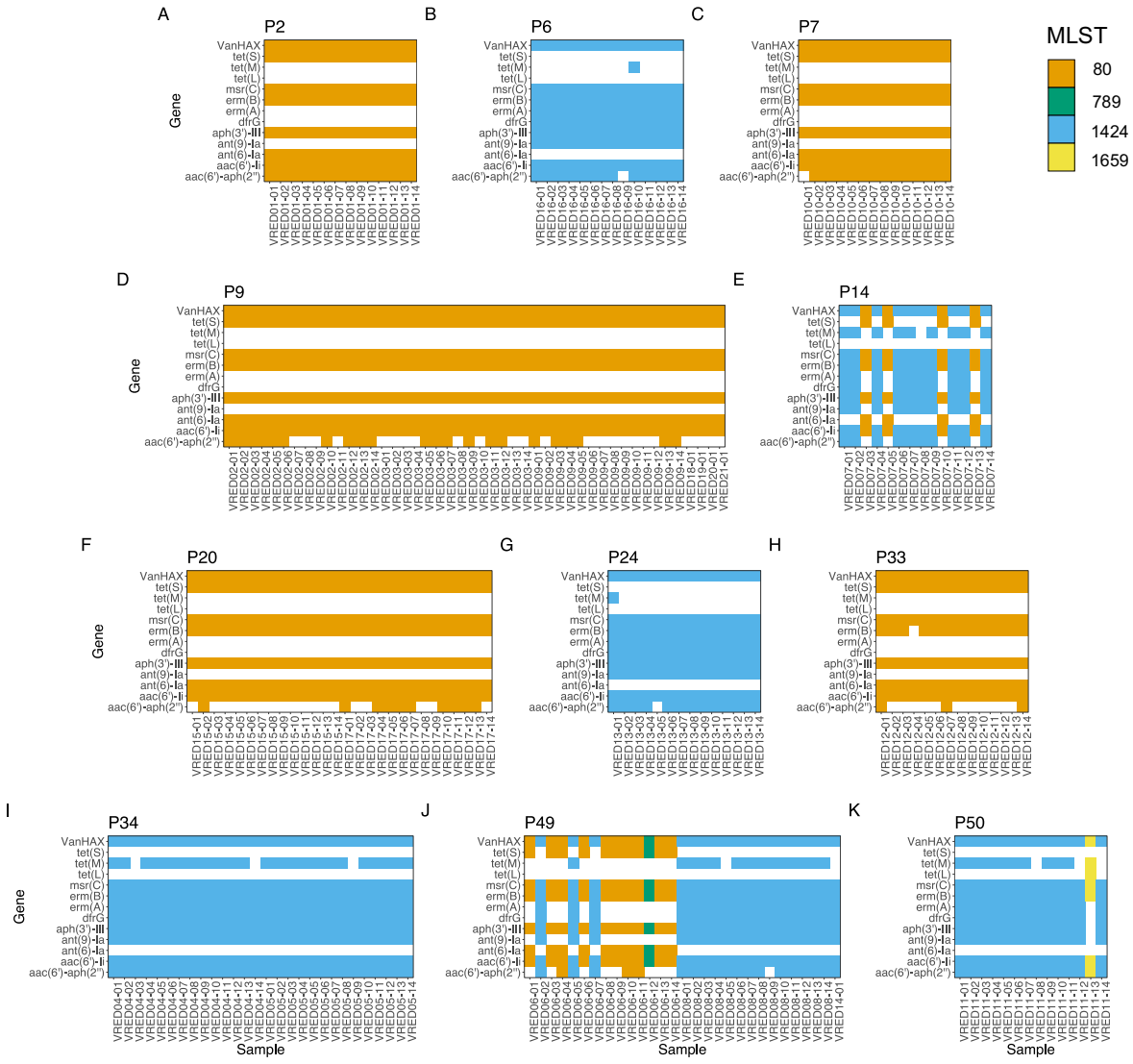


Figure 4.7 Presence of AMR genes varies within patients

Panels A-K represent different patients; resistance genes are plotted on the y-axis and isolates on the x-axis. Presence of a gene in each of the colonies sequenced from that individual is represented by a filled square and coloured based on the MLST of the genome. The absence of a gene is indicated by a white box.

Tetracycline resistance gene *tet(M)* was identified on the chromosome of VRED06-02 as part of Tn*6944* (Figure 4.8A). *tet(M)* was identified in 62.2% of ST1424 and ST1659 genomes; excision of Tn*6944* was responsible for this variable presence. *aac(6')-aph(2'')* was present on p1_VRED06-02 (ST1424) and p1_VRED06-10 (ST80). *aac(6')-aph(2'')* was not detected in any ST80 genomes that were p1_VRED06-10 negative, although only 59.6% (n=65) of genomes that carried this plasmid also carried *aac(6')-aph(2'')*. In p1_VRED06-02, two copies of *aac(6')-aph(2'')* were surrounded by IS*256*, IS*1216*, and IS*3,* providing multiple mechanisms of excision. In ST1424 *aac(6')-aph(2'')* was detected in 97.8% (n=90) genomes with p1_VRED06-02. Another four ST1424 genomes carried *aac(6')-aph(2'')* but not p1_VRED06-02 (Table 4.9 and Figure 4.7). In p1_VRED06-10, *aac(6')-aph(2'')* was surrounded by two copies of IS*256* similarly to Tn*6218*, although the transposition machinery was missing (Figure 4.8B).[372] Short read assemblies could not resolve the environment of the *aac(6')-aph(2'')* gene, but in three cases *aac(6')-aph(2'')* was co-located with an IS*3* gene suggesting mobilisation to another transposable element. The tetracycline resistance gene *tet(L)* was identified in a single ST1659 genome, the gene was co-located with *tet(M)* on a 30 kb contig that was similar to Tn*6248* from *E. faecium* over ~19 kb (Figure 4.8C).

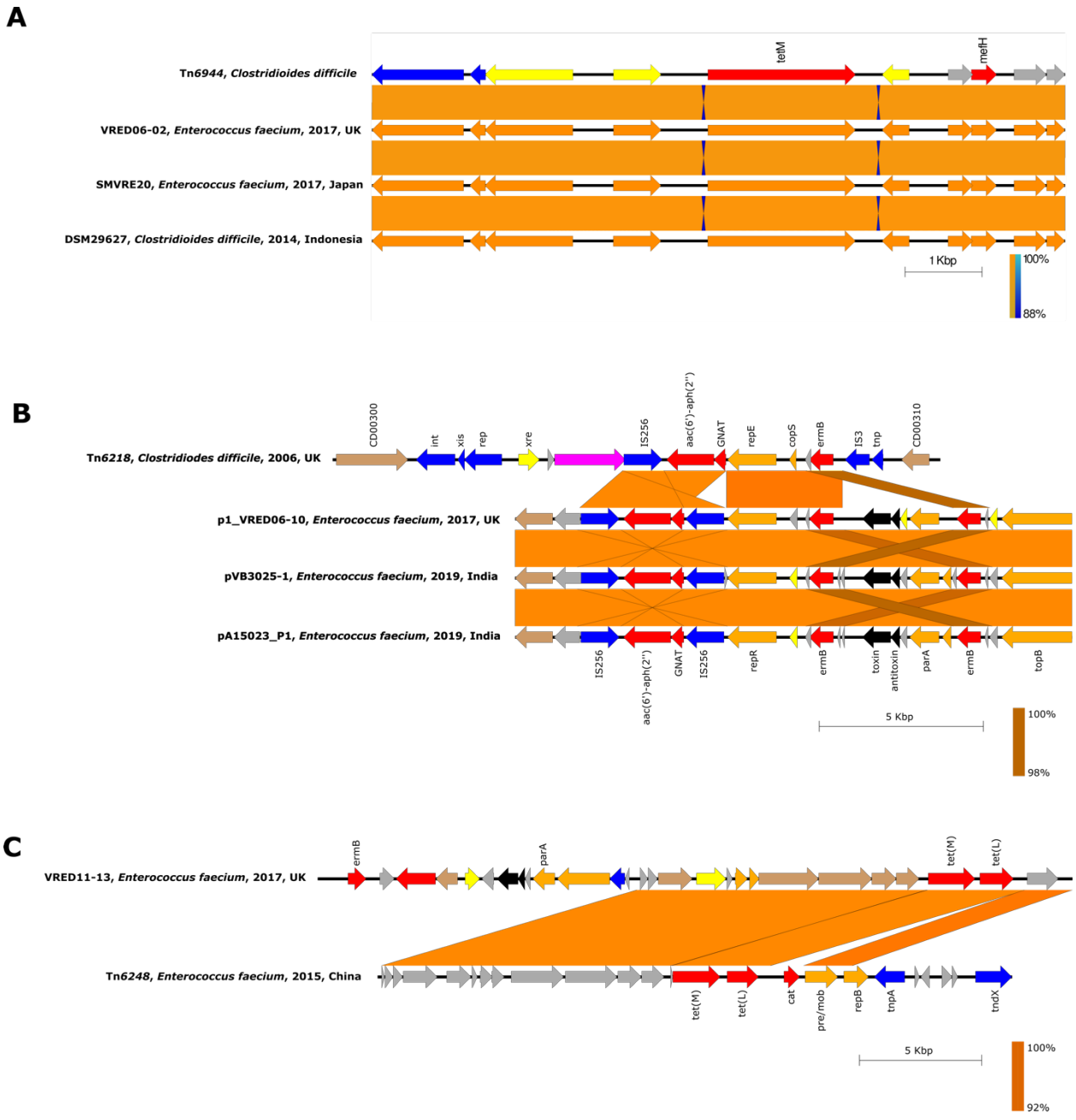


Figure 4.8 Comparison of AMR elements to previously described transposons

Comparison of (A) *tet(M)*, (B) *aac(6')-aph(2'')*, and (C) *tet(L)* elements to published examples. Transposon or strain identification, species, year, and country of first identification are given where available. CDSs are coloured based on inferred function: AMR, red; transposon, blue; replication, orange; regulation, yellow; toxin/antitoxins,

black; hypothetical, grey; pink, surface-associated; other, brown. Coloured blocks between sequences indicate BLASTn identity ≥92%.

4.2.9 Identification of linear plasmid

A single contig in the genome assembly of VRED06-10 did not circularise during assembly and was found to have structural similarity with linear plasmids identified in VREfm from Japan (pELF1, accession LC495616) and Denmark (pV24-2, accession CP036153) (Figure 4.9) [285,373]. The contig in VRED06-10 was named p2_VRED06-02. In pELF1 and pV24-2, the left of the plasmid forms a ~5kb hairpin loop around the 5'-TATA-3' motif and the right hand end contains multiple palindromic sequences that may form hairpins. Proteins are postulated to interact with each end and prevent exonuclease digestion. In p2_VRED06-02, there is a 46kb inverted repeat around the 5'-TATA-3' motif, and a hairpin structure was confirmed. At the right-hand end of p2_VRED06-02, sequence was identical to that of pELF1 and pV24-2. To confirm the contig was linear, PCR primers were designed targeting an internal region within the plasmid as a positive control, and primers targeted off each end (external). A PCR product of expected size was generated from the internal primers, but no product was generated from the external primers, confirming that the identified contig is present within the cell, and is not circularised over the two identified ends (Figure 4.10).

p2_VRED06-10 was 151kb in length, present in a single copy, with 183 CDSs. No *rep* type could be assigned but the *repB* and *parA* genes were identified, which may be used in plasmid replication and segregation [373]. The *vanA* vancomycin resistance operon was the only identified AMR determinant. In total, 136 (74.3%) of the identified CDSs could not be

matched to existing protein databases so these were assigned as hypothetical proteins as a product could not be inferred during annotation. Boumasmoud *et al* [374] recently described the linear plasmid pELF_USZ in VREfm from Switzerland that carried an operon that conferred the ability to utilise the human gut mucin *N*-acetyl-galactosamine, however p2_VRED06-02 did not carry this operon (results not shown).

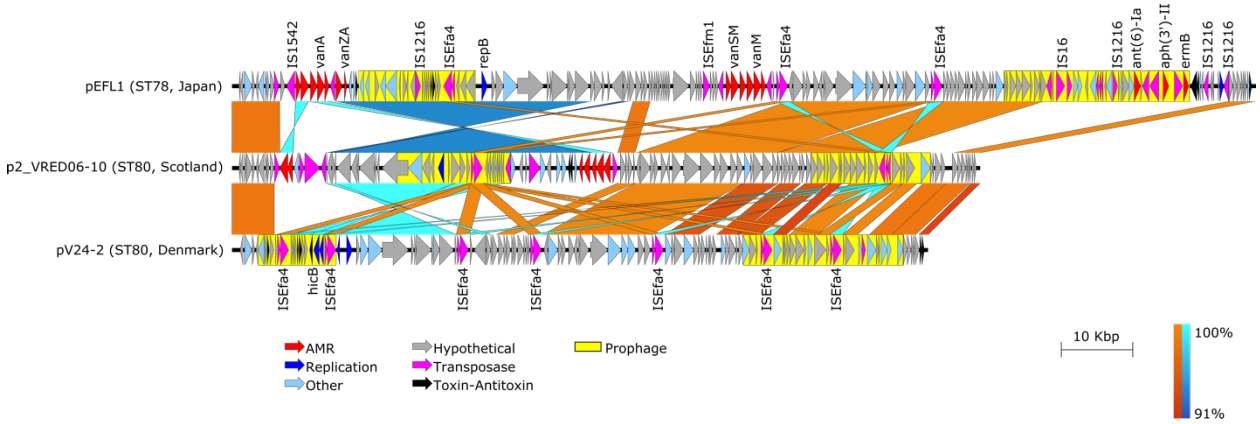


Figure 4.9 Comparison of linear plasmid sequences

Strain identification, ST, and country of first identification are given. CDSs are coloured based on inferred function: AMR, red; hypothetical, grey; replication, dark blue; transposase, pink; toxin/antitoxins, black; prophage, yellow; other, light blue. Coloured blocks between sequences indicate BLASTn identity ≥91%, matches in the same orientation are coloured yellow and inversions are coloured blue. The first half of the inverted repeat has been removed for ease of visualising matched nucleotide blocks; all sequences start at the middle of the hairpin structure (5'-TATA-3') of the inverted repeat.

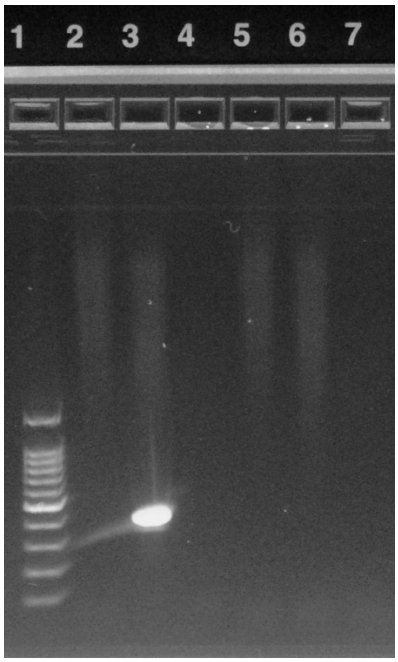


Figure 4.10 – Linear Plasmid PCR products.

Well 1, 100 bp ladder; Well 2, Internal Primer set against VRED06-02; Well 3, Internal Primer set against VRED06-10; Well 4, Internal Primer set against water; Well 5, External Primer set against VRED06-02; Well 6, External Primer set against VRED06-10; Well 7, External Primer set against water.

4.3 Discussion

Within-patient diversity is a potentially important confounder in studies of bacterial transmission [361–367]. Within patient diversity has been described in *E. faecium* but this has not been robustly investigated taking account of statistical power [12,282,368–370]. Published work was utilised to design a sampling strategy to identify the optimal number of colonies to detect transmission and reliably detect within-patient diversity (Tables 4.1 and 4.2). This study performed WGS on 14 colony picks from rectal screening samples collected on a single ward over a one month period. Bacteraemia isolates were included where available to identify how these differed from carriage populations. ST80 and ST1424 dominated the sequenced isolates, and long read sequencing was used to generate high-quality reference genomes within these lineages for short read mapping. Multiple strains were detected in 27% of patients, and bloodstream isolates differed by 0-1 SNP to carriage isolates highlighting transition to invasive disease in this immunosuppressed patient group. The finding of multiple strains in 27% of patients is in line with recent studies showing up to half of patients carry 2-4 different *E. faecium* strains, and within-patient diversity varies over time [267,282,370,375]. This work also identified a maximum of 3 SNPs between isolates of the same lineage within patients, this can be used as a cut-off for transmission based analyses as isolates sharing <3 SNPs can be considered to be very closely related and possibly linked to recent transmission.

Where multiple samples from the same patient were collected over time low (0-3 SNPs) accumulation of SNPs were found with no pattern in the prevalence of other genomic variants. Estimates of diversification rates in *E. faecium* from single colony sampling of national isolate collections suggest 7 mutations per year,[249] other studies of longitudinal within-patient diversification have estimated higher rates of 12.6 – 128 mutations per year.[282,368,369] The low SNP diversity identified in this one-month collection of carriage isolates is in keeping with the lower estimated mutation rates. Sequencing more than 14 colonies would improve the detection of minor variants but would increase costs, complexity, and turnaround time. Given the low within-patient diversity in most patients this approach may not be required in every case. Gouliouris *et al*[376] analysed within-patient diversity of *E. faecium* in 185 stools collected from 109 patients, analysing a median of five (interquartile range 3-5, total 865) colonies. This identified 51% of stools contained multiple *E. faecium* subtypes. Based on the analysis in this thesis five colonies would identify 50% of the population with 95% confidence (Table 4.1), so may be a pragmatic choice for analysing within-patient diversity as it should identify most mixed-strain carriage. However, five colonies gave the lowest confidence in transmission analysis so the accuracy of the linkage method should be considered and verified before applying within-patient diversity estimates to routine transmission investigations in future (Table 4.6, Figure 4.6). Alternatively, strain-resolved metagenomics directly on clinical samples or sweeps of selective culture growth may be more feasible.[377–379] Although these approaches would eliminate considerations over how many colonies to include, they will likely need higher sequencing depth to identify minor variants which would in turn reduce

the number of samples that can be sequenced per run and increase costs. Few analysis pipelines exist for analysing sequence data from mixed samples for transmission analysis, and the suitability of these tools for outbreak investigations remains to be established. Further work is required to determine the optimum sampling strategy to support IPC investigations in healthcare settings.

As within-patient diversity was identified in SNPs as well as MGEs, AMR patterns within the genomes were also investigated to determine if there could be clinically significant variation in susceptibility between single colonies. It was identified that presence of *tet(M)* was variable within individual patients and phenotypic tetracycline susceptibility would therefore differ based on which colony was picked (Figure 4.7). However, tetracyclines are not generally used for treatment of enterococcal human infections so the clinical impact may be limited. Similar variable presence of the vancomycin resistance element within patients has been described elsewhere and could lead to inappropriate use of vancomycin when the patient harbours a resistant subpopulation.[12,282,380,381] This study only included vancomycin resistant isolates, so cannot resolve the potential role of variable vancomycin resistance carriage within patients or in transmission networks.[382] Gain and loss of vancomycin resistance has been described in regional networks over periods of years[249]. Similar variability was identified with the *aac(6')-aph(2")* aminoglycoside resistance gene. The impact on aminoglycoside resistance phenotype is unclear – all genomes carried *aac(6')-Ii* and *aph(3')-III* which together confer high-level resistance to the clinically relevant aminoglycosides amikacin and gentamicin, so the loss

of *aac(6')-aph(2'')* may be more efficient for the cell without an overt change in antibiotic susceptibility. *tet(M)* and *aac(6')-aph(2'')* were present on transposons Tn*6944* and Tn*6218* respectively, both of which were first characterised in *Clostridioides difficile*, highlighting transmission of AMR elements between nosocomial pathogens as recently described.[383]

Long read sequencing allowed the identification of a linear plasmid, p2_VRED06-10, which carried the VanA vancomycin resistance operon. Linear plasmids are increasingly identified in VREfm in studies using long read approaches, before now these would be missed in fragmented short read assemblies [230,285,373,374,384,385]. Many of the genes on p2_VRED06-10 could not be assigned a function from automated annotation although other studies have identified AMR genes and biosynthetic clusters linked to nutrient acquisition. Linear plasmids in enterococci appear to be structurally conserved in a pELF1-like family and are globally distributed, they are highly stable within *E. faecium* due to high horizontal transmission rates, low-level transcription of carried genes, low impact on chromosomal transcription, and a low overall fitness cost [230]. Further investigation of the contents, maintenance, and transfer of enterococcal linear plasmids will be an important aspect of genomic surveillance in the future.

This study has some limitations. Around 60% of *E. faecium* carriers can be linked to nosocomial transmission from other patients or reservoirs in the hospital environment.[125,375,386–388] This study did not include environmental samples, and although

patients were mostly located in individual rooms bathroom facilities were shared posing a significant environmental reservoir for VREfm. Also, direct plating to solid VREfm screening agar was used to identify carriers for inclusion in the study. Previous studies have shown a sensitivity of 58-96% for this approach, rising to 97-100% with a pre-enrichment step.[389–391]

A proactive sequence-based surveillance approach should avoid large infection outbreaks, and reduce ward closure costs and the clinical impact of invasive disease.[392–395] Recent impact modelling for the UK estimated routine WGS-based surveillance could prevent 74,408 HAIs and 1257 deaths while saving £478 million, or £7.83 per £1 invested [396]. In the study setting, an outbreak of VREfm was suspected three weeks after the study collection period when P9 and P49 developed BSI concurrently but this was many weeks after VREfm transmission had likely occurred (Figure 4.1 and Figure 4.4). Prospective WGS surveillance would have highlighted linked cases prior to the increase in BSIs. Due to this study's retrospective nature, sequencing results could not be used to directly influence patient care.

To conclude, by taking account of within-patient diversity in VREfm carriage populations transmission links were identified between patients that could supplement efforts to control transmission within hospitals. This study also show that diversity exists not just at the level of SNPs – AMR gene presence/absence, indels, and plasmid presence all vary within and between patients. Accounting for within-patient diversity is important for

resolving VREfm transmission using WGS-based investigations and therefore its potential to informing infection prevention control measures and control the spread of VREfm.

Chapter 5 Whole genome sequencing based investigation of a suspected nosocomial outbreak of vancomycin resistant *Enterococcus faecium*

5.1 Introduction

Having investigated within-patient diversity and defining a 3 SNP cut-off for putative transmission links, this was then taken forward and applied to a known VREfm nosocomial outbreak. Outbreak investigations are often initiated in response to an increase in VREfm infections or colonisation detected from screening, with the aim being to identify whether transmission has occurred in the clinical setting and stop further transmission. VREfm can be carried in the gut with no symptoms, this is a significant risk for onward transmission in hospital settings and infection control policies aim to limit this [397,398]. Outbreak investigations are often complex requiring input from specialist staff, closure of beds or entire wards, an increase in testing, and enhanced cleaning leading to costs of $159k - $357k per VREfm outbreak [127,399,400].

Typing in outbreak investigations has historically been performed with PFGE but WGS is increasingly used to identify genetic relationships with higher resolution [150,249,375,394]. The analysis that supports this has mainly been based on the definition of a core genome specific to the analysis and detection of SNPs, usually after mapping to a reference genome [254,260]. An alternative approach is cgMLST which compares 1,400 genes common in *E. faecium* between all genomes in the investigation [198]. Core SNP and cgMLST analyses

are relatively stable but ignore a significant amount of genetic diversity, up to 40% of the *E. faecium* genome is estimated to be MGEs much of which will not be included in core genome approaches [252]. Lately, reference-free SKA typing has been shown to be equally or more sensitive for *E. faecium* genetic relationships than both core SNPs and cgMLST [305,401]. SKA uses split *k*-mers to identify variation between closely related genomes, calculate pairwise distances, and create clusters of linked genomes; the use of *k*-mers also makes SKA faster to run than other genomic epidemiology tools [187]. SKA does not rely on a reference genome and so takes account of the whole genome, including MGEs, so should allow more in-depth comparisons between genomes by including more sequence variants [187]. Recently, SKA has been implemented into the PopPIPE pipeline. First, related genomes are clustered using PopPUNK, then PopPIPE calculates the core and accessory distances within each cluster, runs SKA to generate within-cluster alignments, builds a phylogeny for each cluster, and the phylogeny is partitioned into subclusters using fastbaps [185,402]. PopPIPE provides a straightforward means to generate SKA-based subclusters for identifying closely related genomes in a large collection.

In Chapter 4, within-patient diversity was investigated and a maximum of 3 SNPs were identified between related isolates within patients. This knowledge was next applied to investigate a suspected nosocomial outbreak as these are important drivers of transmission of MDR enterococci and are challenging to manage. VREfm were isolated as part of a suspected multi-ward outbreak on the orthopaedic department of RIE in 2016, only single colonies were available for most samples as this is the standard practice

currently. The aim of the study was to investigate the utility of merged WGS and epidemiological analysis to understand suspected VREfm outbreaks. Reference-free and core SNP based clustering will be compared to each other and to PFGE, and the linkage of these genetic clustering approaches with epidemiology will be investigated.

5.2 Results

5.2.1 Epidemiological context

An outbreak investigation was started when a patient (patient 3) on Ward B in the Orthopaedic Department developed VREfm UTI after contact with a known carrier (patient 1). IPC measures were implemented including enhanced cleaning of ward, practice reviews, staff education, and patient screening for VREfm carriage. On day 11 a further investigation was implemented on Ward A when two patients (patients 2 and 11) in a shared room developed VREfm UTI. On day 26 the investigation was further widened when VREfm was isolated from urine and deep tissue samples collected on wards C and D (patients 23 and 24). The investigation was further expanded to Ward E on day 49 when VREfm was isolated from a hip fluid aspirate (patient 42). Patient 42 was the last associated with confirmed clinical infection, screening was discontinued on day 111, and the incident was closed on day 155.

In total, 2114 samples from 1519 patients were investigated for the presence of VREfm, of which 108 (5.1%) samples from 87 (5.7%) patients were positive. Of these, 87 isolates from 84 patients were available for further investigation (Table 5.1). The outbreak investigation followed local guidance at the time and so single colonies were sent for typing and stored in most cases, only the stored colonies were available for this study. While this does not reliably distinguish within-patient diversity, it does reflect current NHS practice and so the findings from this study can be directly applied to other VREfm outbreak investigations. VREfm was isolated throughout the investigation on all implicated wards, with no clear pattern identifiable from epidemiological data alone (Figure 5.1).

Table 5.1 Patient demographics (84 patients)

Metric		**Number (%)**
Female		51 (60.7)
Age, median (range) years		78 (33-98)
Patient Status [a]	Colonised	74 (88.1)
	Possible Infection	7 (8.3)
	Confirmed Infection	3 (3.6)
Location	Ward A	24 (28.6)
	Ward B	23 (25.0)
	Ward C	10 (11.9)
	Ward D	13 (15.5)
	Ward E	12 (14.3)
	ADM	4 (4.8)
Specimen type	Rectal Swab	50 (59.5)
	Faeces	14 (16.7)

	Swab, site not stated	9 (10.7)
	Mid-Stream Urine	5 (6)
	Catheter specimen urine	2 (2.4)
	Perineal Swab	2 (2.4)
	Hip Fluid	1 (1.2)
	Tissue	1 (1.2)

[a] Colonised, VREfm isolated from rectal carriage sample only; possible infection, VREfm isolated from clinical sample in absence of clinical symptoms; confirmed infection, VREfm isolated from clinical sample with symptoms.

ADM, pre-admission clinic

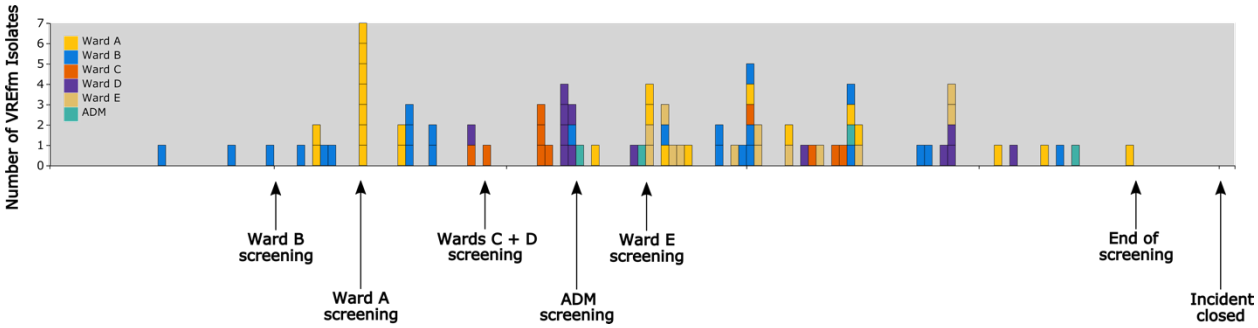


Figure 5.1 Epicurve of VREfm outbreak on Orthopaedic Department

Blocks indicate collection of a VREfm positive sample, blocks are coloured according to patient ward at time of sample collection, time unit is days, incident measures are indicated by arrows. ADM, pre-admission clinic.

5.2.2 Genomic clustering

The 87 genomes in the collection were assigned to five STs, all belonging to the Clade A nosocomial lineage (Table 5.2) [371]. ST80 accounted for 58.6% (n=51) of the genomes, ST203 for 14.9% (n=13), ST18 and ST262 for 12.6% (n=11) each, and ST2287 for 1.1% (n=1). ST2287 was a novel ST, and a single locus variant of ST17.

Core genome SNP analysis grouped 74 (85.1%) genomes into 14 clusters with a median size of 4 (Figure 5.2, Table 5.2). PopPUNK partitioned the genomes into seven clusters, these were like the ST groupings identified by MLST, but PopPUNK split the ST80 and ST203 populations. The same split was seen in the structure of the core SNP phylogeny (Figure 5.2). PopPIPE was used to subcluster within each PopPUNK cluster and assigned 77 (88.6%) genomes into 20 clusters with median size of 3 (Table 5.2). Core SNP and PopPIPE grouping agreed in 68 (78.2%) genomes. Core SNP included three genomes in clusters that PopPIPE identified as singletons, whereas PopPIPE included six genomes in clusters that were singletons with core SNP analysis. PFGE was performed on 84 isolates and grouped 58 (69.1%) into 11 clusters with median size of 4 (Table 5.2). PFGE agreed with core SNP clustering in 52 (61.9%) and with PopPIPE in 50 (59.5%) cases.

Table 5.2 Typing results of VREfm (n = 87)

MLST		PFGE [a]			Core SNP			PopPIPE			Study Wards	Days between first and last isolate
Sequence Type	Isolates	Clusters	Isolates clustered (%)	Median isolates per cluster (Min-max)	Clusters	Isolates clustered (%)	Median isolates per cluster (Min-max)	Clusters	Isolates clustered (%)	Median isolates per cluster (Min-max)		
80	51	6	37 (74)	4 (2-15)	8	43 (84.4)	3 (2-12)	11	48 (94.2)	3 (2-11)	6	116
203	13	3	9 (75)	3 (2-4)	3	12 (92.4)	4 (2-6)	4	11 (84.7)	2.5 (2-4)	5	116
18	11	1	5 (50)	5 (5-5)	2	9 (81.9)	4.5 (4-5)	3	10 (91)	3 (2-5)	6	53
262	11	1	7 (63.7)	7 (7-7)	1	10 (91)	10 (10-10)	2	8 (72.8)	4 (3-5)	3	82
2287	1	0	-	-	0	-	-	0	-	-	1	-
All	87	11	58 (69.1)	4 (2-15)	14	74 (85.1)	4 (2-12)	20	77 (88.6)	3 (2-11)	6	125

[a] 84 isolates tested with PFGE

MLST, multilocus sequence typing; PFGE, pulsed field gel electrophoresis; SNP, single nucleotide polymorphism

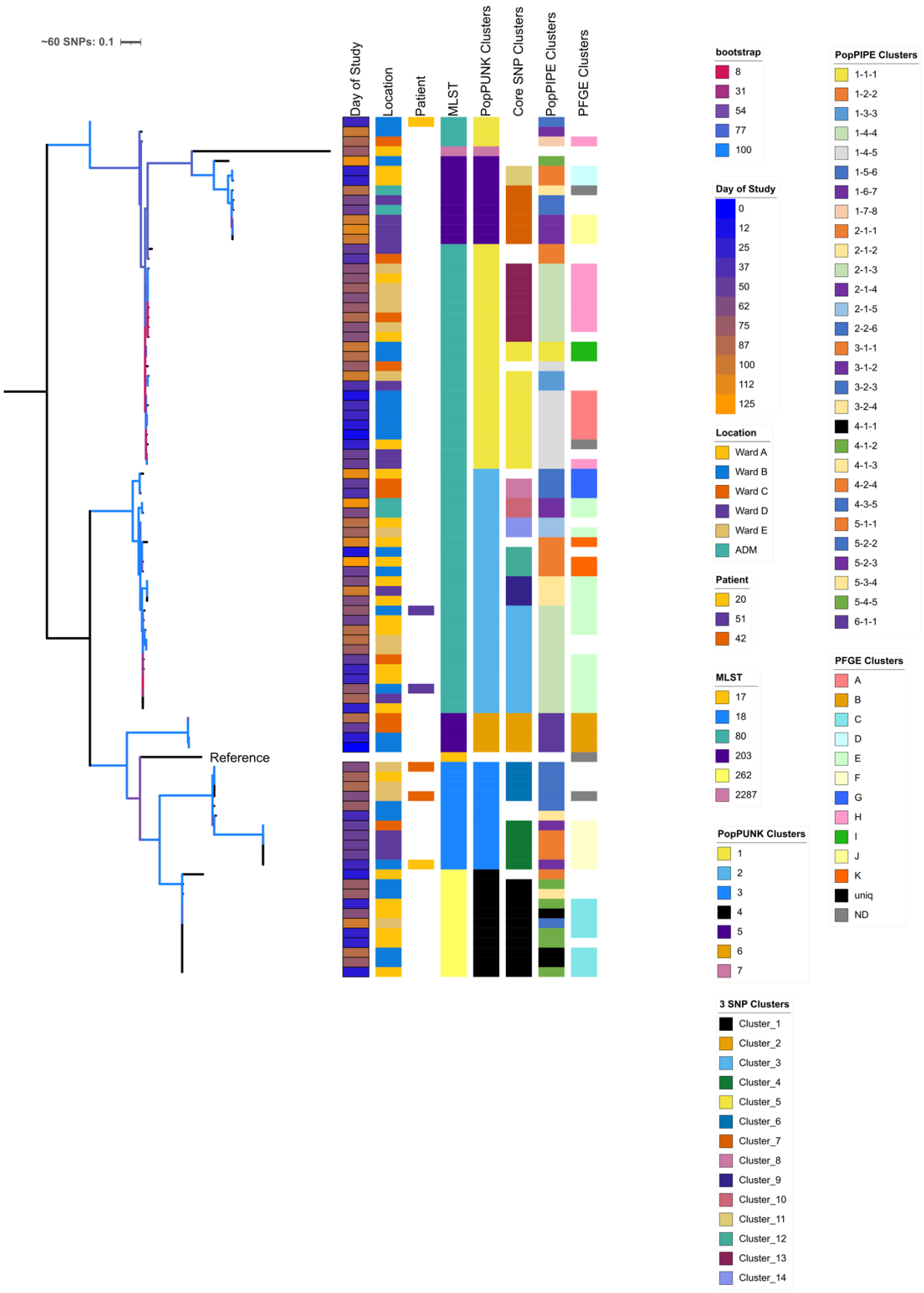

~60 SNPs: 0.1
Reference
Day of Study
Location
Patient
MLST
PopPUNK Clusters
Core SNP Clusters
PopPIPE Clusters
PFGE Clusters
bootstrap
8
31
54
77
100
Day of Study
0
12
25
37
50
62
75
87
100
112
125
Location
Ward A
Ward B
Ward C
Ward D
Ward E
ADM
Patient
20
51
42
MLST
17
18
80
203
262
2287
PopPUNK Clusters
1
2
3
4
5
6
7
3 SNP Clusters
Cluster_1
Cluster_2
Cluster_3
Cluster_4
Cluster_5
Cluster_6
Cluster_7
Cluster_8
Cluster_9
Cluster_10
Cluster_11
Cluster_12
Cluster_13
Cluster_14
PopPIPE Clusters
1-1-1
1-2-2
1-3-3
1-4-4
1-4-5
1-5-6
1-6-7
1-7-8
2-1-1
2-1-2
2-1-3
2-1-4
2-1-5
2-2-6
3-1-1
3-1-2
3-2-3
3-2-4
4-1-1
4-1-2
4-1-3
4-2-4
4-3-5
5-1-1
5-2-2
5-2-3
5-3-4
5-4-5
6-1-1
PFGE Clusters
A
B
C
D
E
F
G
H
I
J
K
uniq
ND

Figure 5.2 Core SNP phylogeny of outbreak collection

All genomes (n=87) mapped to Aus0004 chromosome (2,955,294 bp), MGEs and recombination masked (456,461 bp), and maximum likelihood phylogeny built on remaining SNPs (607 bp). Patient metadata and clustering are indicated by coloured blocks (see figure for key), branches are coloured according to bootstrap. Patient 20, 42, and 51 had multiple isolates sequenced and are indicated in the tree based on coloured blocks in the patient column. Tree is midpoint rooted.

5.2.3 Epidemiological support for genomic clusters

Next, epidemiological linkage was assigned to all patient pairs within the identified clusters to determine to what degree the genomic linkage was supported by patient movement in time and space. It was considered that patients sharing time on the same ward were epidemiologically linked and represent possible direct transmissions, patients sharing the same ward without overlapping stay or having overlapping stay on different wards were also epidemiologically linked and may represent indirect transmission, patients admitted to different wards at different times were weakly linked and transmission opportunity was unclear, all other patients were considered unlinked epidemiologically and transmission could be ruled out.

MLST clustering was not well supported by epidemiological linkage with 9% of pairs admitted to the same ward at the same time, 74% admitted to different wards, and 6% had no identified link (Table 5.3). This reflects the broad clustering of MLST based on long-term evolutionary relationships rather than short-term genetic linkage. Patient pairs clustered with core SNPs had higher agreement with epidemiological linkage, with 26% being on the same ward at the same time and 5% having no identified epidemiological link (Table 5.3). PopPIPE performed similarly but had slightly higher agreement with epidemiological linkage, with 30% of pairs being on the same ward at the same time and 3% having no identified link (Table 5.3). PFGE had similar epidemiological linkage to the

core SNP and PopPIPE clusters, although it had the lowest proportion with no linkage and the highest with unclear linkage (different wards at different times, Table 5.3).

Table 5.3 Epidemiological linkage within genomic clusters

Epidemiological linkage	MLST [n=1354]	PFGE [n=185]	Core SNP [n=228]	PopPIPE [n=159]
	% (95% CI)	% (95% CI)	% (95% CI)	% (95% CI)
Same ward, same time	9.1 (7.6-10.6)	25.5 (19.2-31.8)	25.9 (20.2-31.6)	30.2 (23.1-37.3)
Same ward, stay 1-28d apart	11.7 (10-13.4)	11.9 (7.2-16.6)	14.1 (9.6-18.6)	12.6 (7.4-17.8)
Different ward, same time	30.0 (27.6-32.4)	25.5 (19.2-31.8)	22.0 (16.6-27.4)	26.5 (19.6-33.4)
Different ward, stay 1-28d apart	43.5 (40.9-46.1)	36.3 (29.4-43.2)	32.9 (26.8-39.0)	27.7 (20.7-34.7)
No link	5.9 (4.6-7.2)	1.1 (0.0-2.6)	5.3 (2.4-8.2)	3.2 (0.5-5.9)

CI, confidence interval; MLST, multilocus sequence typing; PFGE, pulsed field gel electrophoresis; SNP, single nucleotide polymorphism

Coloured from low (blue) to high (orange) percentage across each row

5.2.4 Cluster introductions and on-ward transmission

Next, patients were assigned as likely VREfm introductions and acquisitions based on timing of positivity relative to admission. Twenty five (29.8%) patients were identified as likely VREfm introductions, three (3.6%) likely acquired VREfm during inpatient episode, and 56 (66.7%) were inconclusive. As PopPIPE clustered the most genomes together, PopPIPE clusters containing likely VREfm introductions were then investigated to identify possible on-ward transmissions. The 25 introductions were present in 21 PopPIPE clusters containing a total of 63 patients (Table 5.4). All three identified acquisitions were VREfm positive after a cluster introduction and were on the same ward at the same time, suggesting these three acquisitions were due to direct transmission from an introduction. There were six introduction cases assigned as singletons by PopPIPE, suggesting these introductions did not lead to any direct transmission in the study population. Of 35 inconclusive cases, two (5.7%) were on the same ward at the same time as an introduction case and may be instances of direct transmission.

Table 5.4 Introductions of VREfm and possible onward transmission

PopPIPE Cluster	Total Patients	Introductions (n)	Acquisitions			Inconclusive		
			n	Same ward as introduction	Same time as introduction	n	Same ward as introduction	Same time as introduction
2-1-3	10	2	1	1	1	7	0	1
1-4-5	9	1	0	-	-	8	0	0
4-1-2	5	1	0	-	-	4	0	0
2-1-1	4	1	1	1	1	2	0	0
3-2-3	4	1	0	-	-	3	0	1
6-1-1	4	2	0	-	-	2	0	0
2-2-6	3	1	0	-	-	2	0	0
4-1-1	3	1	0	-	-	2	0	2
2-1-2	3	2	0	-	-	1	1	1
1-1-1	2	1	1	1	1	0	-	-
1-3-3	2	1	0	-	-	1	1	0
5-2-2	2	1	0	-	-	1	1	1
3-1-2	2	1	0	-	-	1	0	1
2-1-5	2	1	0	-	-	1	0	0
2-1-4	2	2	0	-	-	0	-	-
1-6-7	1	1	0	-	-	0	-	-
1-7-8	1	1	0	-	-	0	-	-
4-1-3	1	1	0	-	-	0	-	-
5-3-4	1	1	0	-	-	0	-	-
5-4-5	1	1	0	-	-	0	-	-
No ID	1	1	0	-	-	0	-	-

Total	**63**	**25**	**3**	**3 (100)**	**3 (100)**	**35**	**3 (8.6)**	**7 (20.0)**

ID, identification

The three largest PopPIPE clusters also had the most inconclusive cases based on timing of positivity, epidemiological data was investigated to attempt to resolve these clusters (Figure 5.3). Cluster 2-1-3 contained 10 patients of which two were introductions, one acquisition, and seven inconclusive based on timing of positivity. Patients 14, 17, 18, and 39 were all on Ward A at the same time so may represent a transmission cluster (Figure 5.3A). While patient 39 was on Ward A there was overlap with patient 28 on Ward C and Patient 51 on non-study wards and Ward B, this may represent between-ward transmission via unidentified sources or detection of pre-existing carriage of related isolates. Patient 60 was identified as an introduction into Ward E, with patient 75 subsequently acquiring VREfm after sharing time on Ward E, this indicates likely transmission on the ward. Patient 67 overlapped in time with patients 60 and 75 but on different wards, they had also stayed on Ward A but prior to any cases in this cluster. Patient 73 screened VREfm positive while a day case on Ward A so was classed as an introduction, they were subsequently admitted to Ward B but no further cases were identified on this ward.

Cluster 1-4-5 contained nine patients of which one was an introduction and eight inconclusive (Figure 5.3B). Patient 1 was a known VREfm carrier, patient 3 grew VREfm from a urine sample in month 3 having shared a room with patient 1 and an outbreak investigation was called. Patient 1 was screened while in a neighbouring hospital to be included in the investigation. Transmission between patient 1 and 3 can be ruled out as the isolate from patient three was assigned to cluster 6-1-1. However, WGS does link

patient 1 and patient 4 (a rectal screen performed as part of the outbreak investigation) and these patients shared time on Ward B suggesting a transmission link. Patients 8, 19, and 21 were subsequently positive on Ward B with overlap on the ward in a likely transmission cluster. Patient 16 was positive on Ward A at the same time. Patients 30 and 34 were positive on Ward D around 2 weeks after these cases had been discharged from the hospital, in a likely on-ward transmission pair. Patient 34 shared a short time in the hospital with patients 16, 19, and 21, and was on Ward D for a month while patient 1 was on Ward A. The introduction case in this cluster was paradoxically the last identified, screening positive at a pre-admission clinic around two weeks after patient 34. Patient 57 had no identified hospital contact prior to screening positive and was admitted from their home.

Cluster 4-1-2 contained five patients of which one was an introduction and four inconclusive (Figure 5.3C). Patients 2 and 11 both had VREfm UTI after sharing a room on Ward A which prompted an outbreak investigation, WGS confirmed these cases are likely transmission but also identified patient 13 and 15 as being linked and on the ward at the same time. Similar to cluster 1-4-5, patient 55 was admitted three weeks later to Ward B and was the only identified introduction of this cluster. Patient 55 had no identified hospital contact prior to screening positive and was admitted from their home.

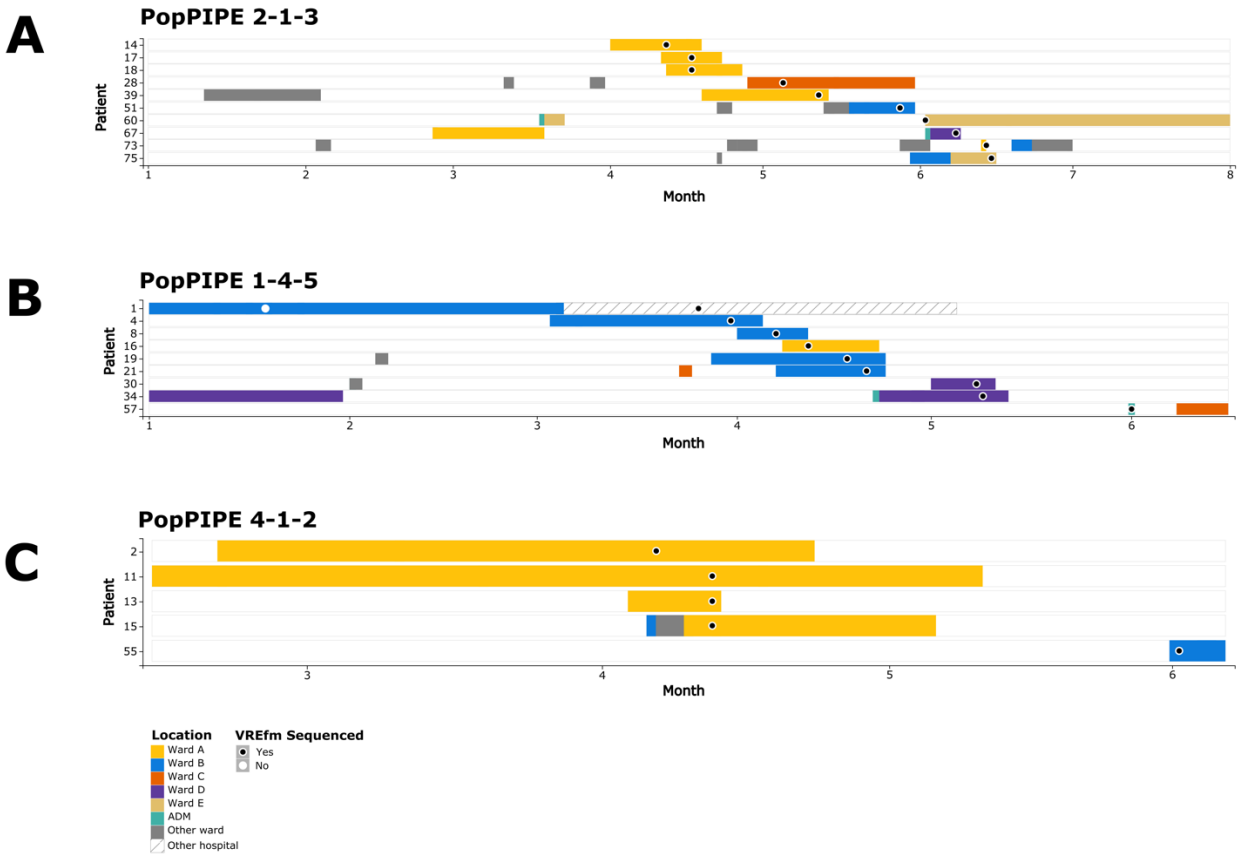


Figure 5.3 Patient timelines for three largest PopPIPE clusters

A, PopPIPE cluster 2-1-3; B, PopPIPE cluster 1-4-5; C, PopPIPE cluster 4-1-2. Each patient is represented by a row, with time on the x-axis. Hospital visits are plotted as coloured boxes, coloured by the ward, isolates are represented by circles and coloured based on whether the isolate was sequenced or not. ADM, pre-admission clinic

5.3 Discussion

The utility of genomic typing for investigating a suspected multi-ward nosocomial outbreak of VREfm was investigated. Among a collection of 87 isolates, core SNPs clustered 74 (85.1%) into 14 clusters and PopPIPE clustered 77 (88.6%) into 20 clusters (Table 5.2). Clustering agreed between the two methods for 68 (78.2%) isolates. Considering epidemiological linkage of patients, 30.2% of patient-pairs in PopPIPE clusters were on the same ward at the same time, compared to 25.9% of pairs in core SNP clusters (Table 5.3).

For both core SNP and PopPIPE clustering, ~54% of patient pairs were on different wards either at the same time or within 28 days (Table 5.3). This could be due to clustering of genetically related lineages circulating outside of the studied setting, or undetected sources of transmission linking multiple wards (staff, patients, equipment, or shared facilities) [127,357,403]. The SNP approach in this study clustered based on SNPs in a core genome, the PopPIPE approach first partitioned the genome collection into broad groupings using PopPUNK and then used SKA to perform reference-free whole genome sub-clustering within these groups. PopPIPE considers the whole genome including mobile elements when clustering, this likely provides further discrimination and may explain why more clusters with smaller numbers were identified with PopPIPE compared to core SNPs (Table 5.2). Similar results have recently been described comparing SKA with SNP-based and cgMLST typing highlighting that split *k*-mer based clusters can have higher

epidemiological support [305,401]. MGEs allow *E. faecium* to adapt to the nosocomial environment and drive the emergence of novel clones, including these elements in outbreak investigation could provide important insights into transmission patterns [252,305,401]. However, MGEs may also spread into lineages that are distinct when core SNPs are considered, this may explain why some genomes were unclustered by core SNPs but clustered by PopPIPE [404]. In comparison to core SNP mapping, PopPIPE and other SKA based approaches also benefit from fast run times using *k*-mers and do not require a well curated reference genome to generate accurate clustering. Higgs *et al*[305] recommend SKA as the optimal method for identifying putative transmission links in *E. faecium.*

This Chapter also confirms PFGE has lower resolution than WGS-based typing [254], clustering less isolates into fewer clusters and agreeing with WGS clustering in 60-62% of cases, although when isolates are clustered by PFGE these have reasonable epidemiological support (Table 5.3). PFGE may retain a role in outbreak investigations if WGS is not available, particularly to rule out transmission when isolates do not have the same PFGE profile, although with the significant caveat that some cases of transmission are likely to be missed or incorrectly assigned.

Twenty five (29.8%) patients were identified as likely VREfm colonised at admission, suggesting a significant baseline of VREfm colonisation in the patient cohort and in keeping with estimates from another centre in the UK [101]. Additionally, 25/35 (71.4%) inconclusive cases in eight PopPIPE clusters with introductions were identified as VREfm

positive before the introductions were admitted (Table 5.4), indicating multiple reservoirs of these genetic clusters. Only 7 (8.3%) patients were admitted directly from another hospital or care home, but the patient cohort is elderly (Table 5.1) and 43 (51.2%) had been admitted in the 3 months before the study began so this cohort could be considered high risk for VREfm carriage. Given the high carriage rate detected (29.8%), structured surveillance screening within 48 h of admission, at least weekly during stay, and at discharge would be optimal to correctly assign introductions, acquisitions, and likely transmission sources but can be challenging to implement [405,406]. Of note, eleven PopPIPE clusters either had one isolate, or only introductions, and so there was no evidence for transmission of these clusters from this analysis.

Three (3.6%) cases of likely VREfm acquisition were identified during hospital stay based on conversion from negative to positive rectal screens. All three cases clustered with identified introductions and shared time on the same wards, indicating a likely patient source for transmission. It was not possible to classify 56 (66.7%) patients as likely introduction or acquisition due to the absence of VREfm screening within 48 h of admission, but the addition of WGS clustering allowed the investigation of likely transmission between some individuals (Figure 5.3). However, the detection of very closely linked genomes from patients admitted to different wards at different times is challenging to interpret even with WGS, and has been reported elsewhere in *E. faecium* [249,407]. Recently, Cassone *et al*[408] performed mathematical modelling to define the spread of VRE in their hospital. This showed that VRE had a high tendency towards horizontal

spread between patients in different rooms, or multiple introductions into the hospital of related strains. Despite the high burden of VRE in their hospital, environmental persistence within individual rooms was rare suggesting terminal cleaning was effective. Taken together with the results in this Chapter, it appears that some of the difficulties investigating VRE outbreaks is due to the continual introduction of related strains into different clinical areas likely due to a significant number of admissions being VRE carriers from previous healthcare contact.

This study was mainly performed on single colony picks but included two morphologically distinct isolates from three patients, giving a small insight into within-patient diversity. These isolates were selected for further typing as they had different colony morphology. In two cases, the within-patient isolates clustered together, but in patient 20 the two isolates were assigned to different STs and differed by 167 core SNPs (Figure 5.2). If one isolate from patient 20 had been selected, they would either be assigned a singleton or clustered with other cases. Carriage of multiple *E. faecium* genetic subtypes has been identified in 27-51% of cases, which can complicate accurate delineation of nosocomial transmission events [376]. A limitation of this study is that multiple colonies were not sequenced in all cases and so some relationships may be missed. Currently, the optimal approach to include within-patient diversity in transmission analyses of *E. faecium* remains to be established, but the core SNP threshold of 3 SNPs used here was based on the maximum within-patient diversity identified in Chapter 4. Additionally, accurate assignation of VREfm carriage status can be affected by the laboratory method used: solid

medium has estimated sensitivity for VREfm of 58-96%, compared to 97-100% using enrichment [389–391]. The findings in this Chapter are applicable to current practice and add to increasing evidence for the utility of SKA-based genomic clustering [305,401].

In summary, in a collection of 87 VREfm isolates from 84 patients implicated in a suspected multi-ward transmission outbreak 85-89% of isolates were clustered based on WGS typing. PopPIPE was slightly more discriminatory than core SNP clustering, likely due to the consideration of MGEs. However, when considering epidemiological linkage there was not always a strong relationship within genetic clusters, suggesting reservoirs of VREfm transmission outside of the setting or transmission mechanisms not identified in this analysis. Although three patients were identified as likely acquisitions and at least 29.8% were colonised on admission, it could not be determined whether 66.7% of patients had acquired VREfm during their stay based on timing of samples, further analysis of WGS clustering identified putative transmission links between the inclusive cases. These findings suggest PopPIPE is a suitable method for VREfm clustering for outbreak investigation, but this is reliant on a robust sampling strategy and patient epidemiological data.

Chapter 6 Genomic analysis of national vancomycin resistant *Enterococcus faecium* dynamics

6.1 Introduction

Having investigated the role of nosocomial transmission on VREfm lineages, the role of regional relationships was investigated next. Surveillance of BSIs has identified an increasing number of *E. faecium* in Scotland that are resistant to vancomycin, comparisons to other countries show a similar increasing trend but suggest Scotland has a higher VREfm rate than England, Australia, and most of Europe (Table 1.3) [21,62,65,409]. The exact reasons for this higher rate are unclear, there could be multiple factors involved including a high-risk patient population, ineffective or poorly applied IPC measures, or the presence of particularly invasive strain(s) of VREfm. At a national level recent studies from Australia, Denmark, England, Germany, and the Republic of Ireland have highlighted diverse populations of VREfm with evidence of spread of VREfm clones between and within hospitals [140,257,410–412].

In Scotland, studies in the mid-1990s using PFGE identified clonal spread of *vanB* positive VREfm in Glasgow hospitals, while in Edinburgh a *vanA* positive outbreak was identified with potential inter-hospital transfer within Edinburgh and to a neighbouring region [413,414]. A 2012 outbreak of linezolid and vancomycin resistant *E. faecium* was described by Inkster *et al* [415], PFGE analysis showed transmission within a single ward in Glasgow. More recently, Lemonidis *et al* [284] performed WGS on five VREfm isolated from two hospitals in Lanarkshire, showing

related ST1424 isolates in the two hospitals and phylogenetic clustering with ST1424 from Australia and England. However, Lemonidis *et al* do not report relationships at the SNP level so it is unclear if these genomes represent direct transmission or not. There is currently a lack of understanding of the genetic epidemiology of VREfm in Scotland.

Chapters 4 and 5 investigated VREfm dynamics within closely linked wards within a single hospital, the aim of this Chapter is to identify the genetic background of VREfm across Scotland to determine if VREfm lineages are geographically limited, or if there is evidence of national transmission patterns.

6.2 Results

6.2.1 Description of collection

This study included 326 isolates identified as VREfm in eight of the 14 regional Health Boards providing frontline healthcare in Scotland between 2012 and 2017 (Figure 6.1, Table 6.1).

Table 6.1 Sampling Health Boards for Scottish VREfm

Health Board	**Count (%)**
Lothian	86 (26.4)

Tayside	77 (23.6)
Greater Glasgow & Clyde	45 (13.8)
Fife	43 (13.2)
Grampian	42 (12.9)
Dumfries & Galloway	19 (5.8)
Highland	8 (2.5)
Lanarkshire	6 (1.8)

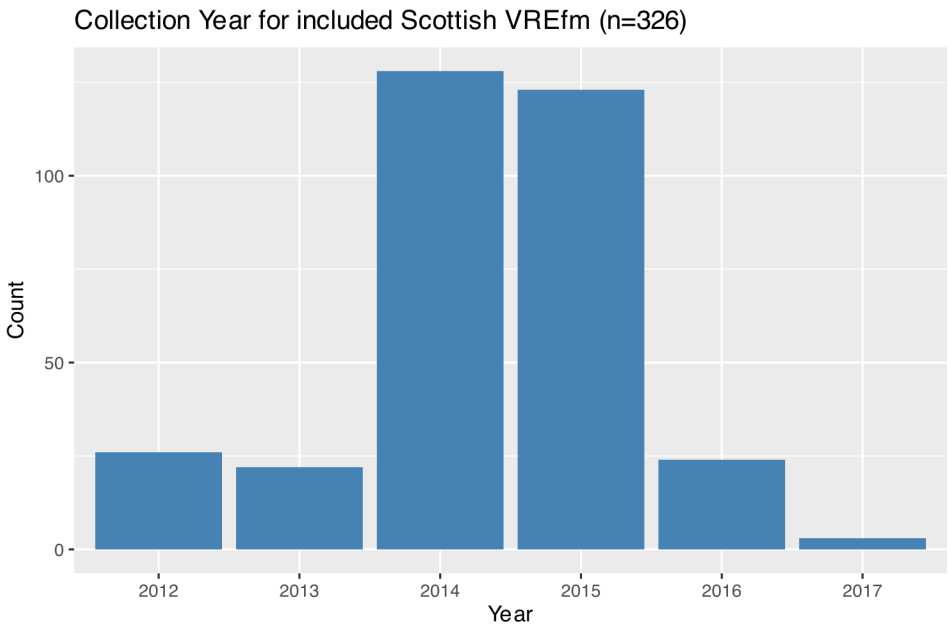


Figure 6.1 Histogram of collection year for the 326 Scottish VREfm isolates

The included isolates represented a convenience sample of all available isolates in participating Health Boards at the time of study inception (2016). This includes isolates stored at the Scottish MRSA Reference Laboratory after PFGE typing as part of outbreak investigations, and isolates stored in regional Health Boards after isolation from clinical or screening samples. Isolates from a range of clinical samples were included, although blood, rectal, and urine samples accounted for 20-24% each (Table 6.2).

Table 6.2 Clinical samples types yielding study isolates

Sample Type	**Count (%)**
Blood	78 (23.9)
Rectal screen	78 (23.9)
Urine	66 (20.2)
NA	53 (16.3)
Drain	23 (7.1)
Tissue/bone	16 (4.9)
Wound	7 (2.1)
Respiratory	3 (0.9)
Line	2 (0.6)

NA, Not available

We estimate that the collection includes ~25% of all bloodstream VREfm from Scotland in 2012-2015 during which time VREfm detections went from stable to increasing year on year (Table 6.3, Figure 1.2).

Table 6.3 Proportion of Bloodstream Scottish VREfm isolates included in this study

Year	Study count	National count[1]	Estimated study coverage of national count (%)
2012	13	61	21.5
2013	14	55	25.6
2014	23	87	26.4
2015	26	102	25.6
2016	2	114	1.8
2017	0	124	0.0

[1]Data from [416,417]

6.2.2 MLST

The Scottish genomes were assigned to 20 STs, all related to the previously described Clade A nosocomial lineage [371]. ST203 and ST80 accounted for 74% of the genomes with other STs accounting for lower numbers of isolates (Table 6.4). All STs with more than one assigned genome were detected in multiple Health Boards, although ST distribution varied between regions. ST203 was particularly concentrated in Tayside and Lothian, ST80 was more common in

Fife, Grampian, Lothian, and Dumfries and Galloway, Greater Glasgow and Clyde had the most STs detected (n=10) and was the most common source of ST262 and ST17.

Table 6.4 MLST Results in Scottish isolates (n=326)

ST	Total (%)	Lothian	Tayside	Greater Glasgow & Clyde	Fife	Grampian	Dumfries & Galloway	Highland	Lanarkshire
203	125 (38.3)	50	54	9	3	6	1	0	2
80	117 (35.9)	21	8	8	34	28	14	3	1
262	23 (7.1)	4	2	11	3	0	0	0	2
18	15 (4.6)	3	9	0	0	0	0	3	0
17	12 (3.7)	2	1	8	0	1	0	0	0
2228	8 (2.5)	0	0	4	0	0	2	2	0
64	7 (2.1)	0	0	0	0	6	1	0	0
412	5 (1.5)	2	0	1	2	0	0	0	0
78	2 (0.6)	0	0	1	0	0	0	0	1
117	2 (0.6)	1	0	0	1	0	0	0	0
132	1 (0.3)	0	0	1	0	0	0	0	0
280	1 (0.3)	0	0	0	0	0	1	0	0
282	1 (0.3)	1	0	0	0	0	0	0	0

992	1 (0.3)	0	0	0	0	1	0	0	0
1032	1 (0.3)	0	0	1	0	0	0	0	0
2227	1 (0.3)	0	0	1	0	0	0	0	0
2229	1 (0.3)	0	1	0	0	0	0	0	0
2230	1 (0.3)	1	0	0	0	0	0	0	0
2231	1 (0.3)	0	1	0	0	0	0	0	0
2232	1 (0.3)	0	1	0	0	0	0	0	0
Total	326 (100)	85	77	45	43	42	19	8	6

ST, Sequence Type

6.2.3 Genomic clustering of Scottish VREfm

The population structure of the Scottish VREfm isolates was investigated by reference based mapping and generation of a SNP-based core genome phylogeny (Figure 6.2). The direct transmission threshold used in Chapter 5 was doubled to 6 SNPs to identify genomes linked in putative regional transmission networks. A total of 238 (73.0%) genomes were assigned to 17 clusters using a threshold of 6 SNPs, which highlighted intra- and inter-regional VREfm spread (Table 6.5). Clusters 5, 6, 7, 10, 13, and 15 contained genomes from single Health Boards and may represent local lineages of transmission either within a single hospital or local healthcare network. Clusters 9 and 11 were the widest spread clusters and were both detected in five Health Boards. Cluster 9 is dominated by genomes from Fife, with smaller numbers of detections in Lothian, Tayside, Highland, and Grampian. Cluster 17 contained 44 genomes from Tayside and 1 from Fife, it is the biggest cluster in the collection. Cluster 14 contains 40 genomes, mainly from Lothian but also three genomes each from Tayside and Grampian. Cluster 14 is not only of interest as a large cluster from Lothian with evidence of transfer to other Health Boards, but also because Clusters 15, 16, and 17 arose from within it suggesting diversification within the ST203 lineage in Scotland.

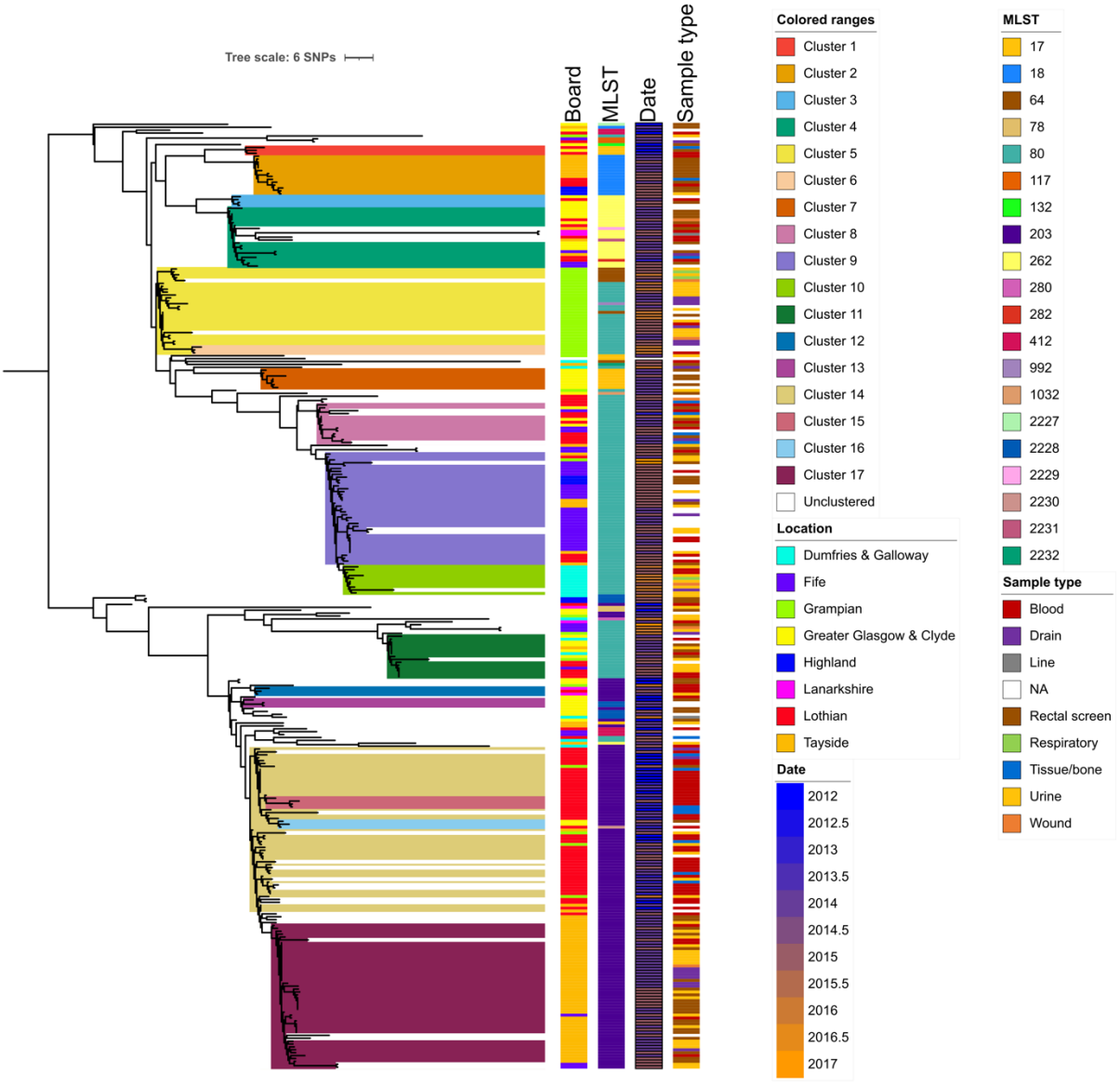


Figure 6.2 Core SNP phylogeny of Scottish VREfm

All genomes (n=326) mapped to Aus0004 chromosome (2,955,294 bp), MGEs and recombination masked (1,994,316 bp), and maximum likelihood phylogeny built on remaining SNPs (1080 bp). Clusters were assigned where three or more genomes had ≤6 SNPs. Metadata and clustering are indicated by coloured blocks (see figure for key).

Table 6.5 VREfm clusters identified by collecting Health Board

Clusters	Lothian	Tayside	Greater Glasgow & Clyde	Fife	Grampian	Dumfries & Galloway	Highland	Lanarkshire	Total (%)
Cluster 1	2	-	1	-	-	-	-	-	3 (1.3)
Cluster 2	3	8	-	-	-	-	3	-	14 (5.9)
Cluster 3	1	-	3	-	-	-	-	-	4 (1.7)
Cluster 4	4	-	8	2	-	-	-	-	14 (5.9)
Cluster 5	-	-	-	-	25	-	-	-	25 (10.5)
Cluster 6	-	-	-	-	3	-	-	-	3 (1.3)
Cluster 7	-	-	7	-	-	-	-	-	7 (2.9)
Cluster 8	6	-	3	2	-	-	-	-	11 (4.6)
Cluster 9	4	6	-	22	1	-	3	-	36 (15.1)
Cluster 10	-	-	-	-	-	9	-	-	9 (3.8)
Cluster 11	5	1	5	1	-	2	-	-	14 (5.9)
Cluster 12	1	-	-	-	-	-	-	2	3 (1.3)
Cluster 13	-	-	3	-	-	-	-	-	3 (1.3)
Cluster 14	34	3	-	-	3	-	-	-	40 (16.8)
Cluster 15	4	-	-	-	-	-	-	-	4 (1.7)
Cluster 16	1	-	2	-	-	-	-	-	3 (1.3)
Cluster 17	-	44	-	1	-	-	-	-	45 (18.9)

6.2.4 Comparison of national collection to known outbreaks

Given the detection of local clusters and evidence for diversification leading to new clusters of transmission in the national collection, the outbreak genomes described in Chapters 4 and 5 were compared to the national collection. All reads were mapped to the Aus0004 reference and a core SNP phylogeny generated (Figure 6.3). This showed that the Chapter 4 and 5 genomes clustered within the lineages of ST18, ST80, ST203, and ST262 identified in Lothian, suggesting these outbreaks were largely driven by ongoing VREfm transmission dynamics within Lothian.

The ST1424 genomes from Chapter 4 were distinct from others in the phylogeny (Figure 6.3). Lemonidis *et al*[284] describe four ST1424 VREfm isolated in Lanarkshire also in 2017, to investigate whether these cases represented a national cluster all Scottish ST1424 were aligned to the V24 reference genome (ST80 isolated in Denmark in 2013) and a maximum likelihood phylogeny generated (Figure 6.4). This showed the Lothian and Lanarkshire ST1424 genomes represented related but distinct populations, separated by 12 SNPs. Within Lanarkshire, 9-30 SNPs differentiated the four genomes suggesting these were not linked by recent direct transmission and are part of a more diverse population.

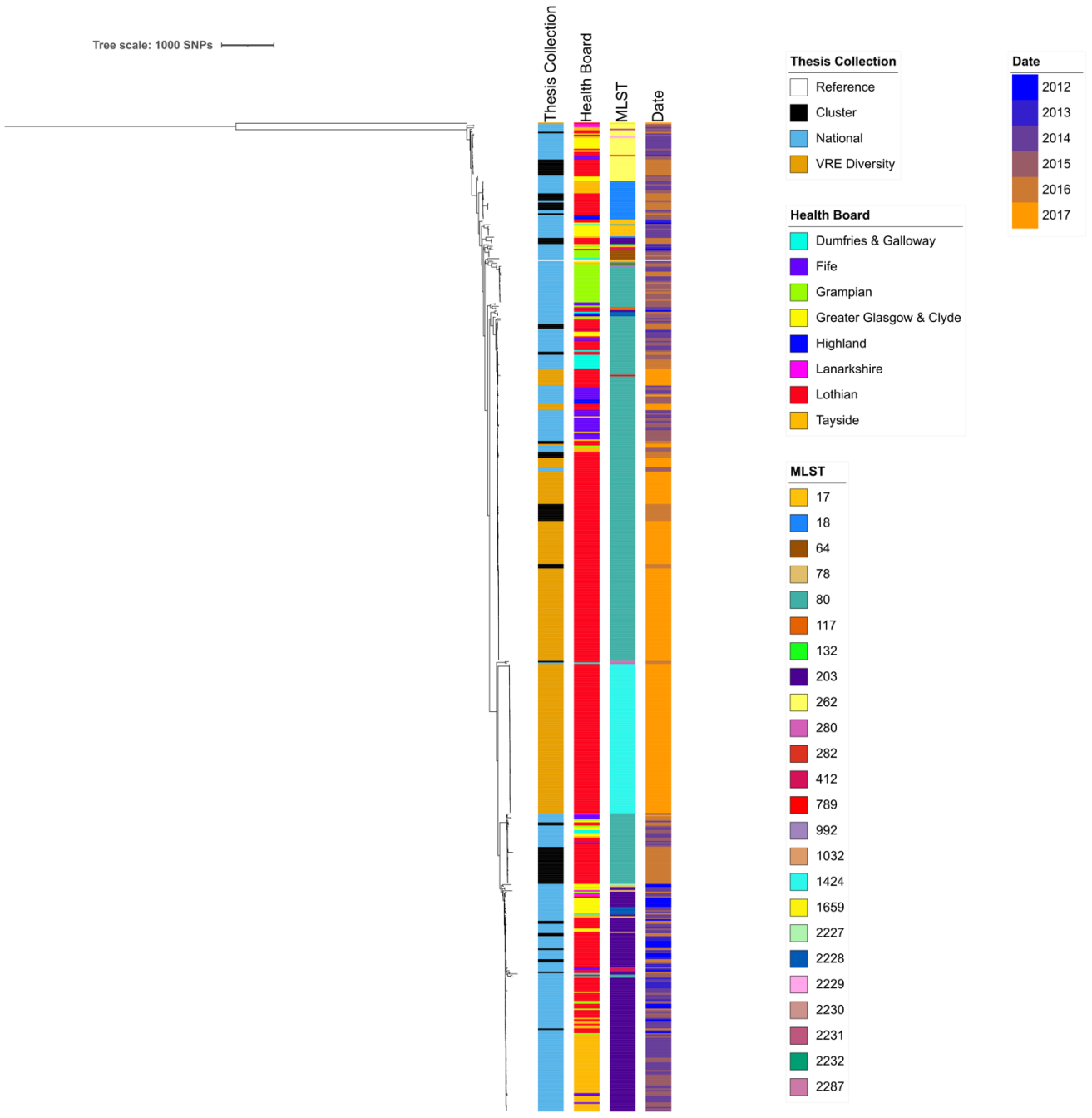


Figure 6.3 Phylogeny of all Scottish VREfm genomes presented in this thesis

Genomes (n=642) mapped to Aus0004 chromosome (2,955,294 bp), MGEs and recombination masked (2,146,548 bp), and maximum likelihood phylogeny built. Metadata

are indicated by coloured blocks (see figure for key). Thesis collection: VRE Diversity, Chapter 4; Cluster, Chapter 5; National, Chapter 6.

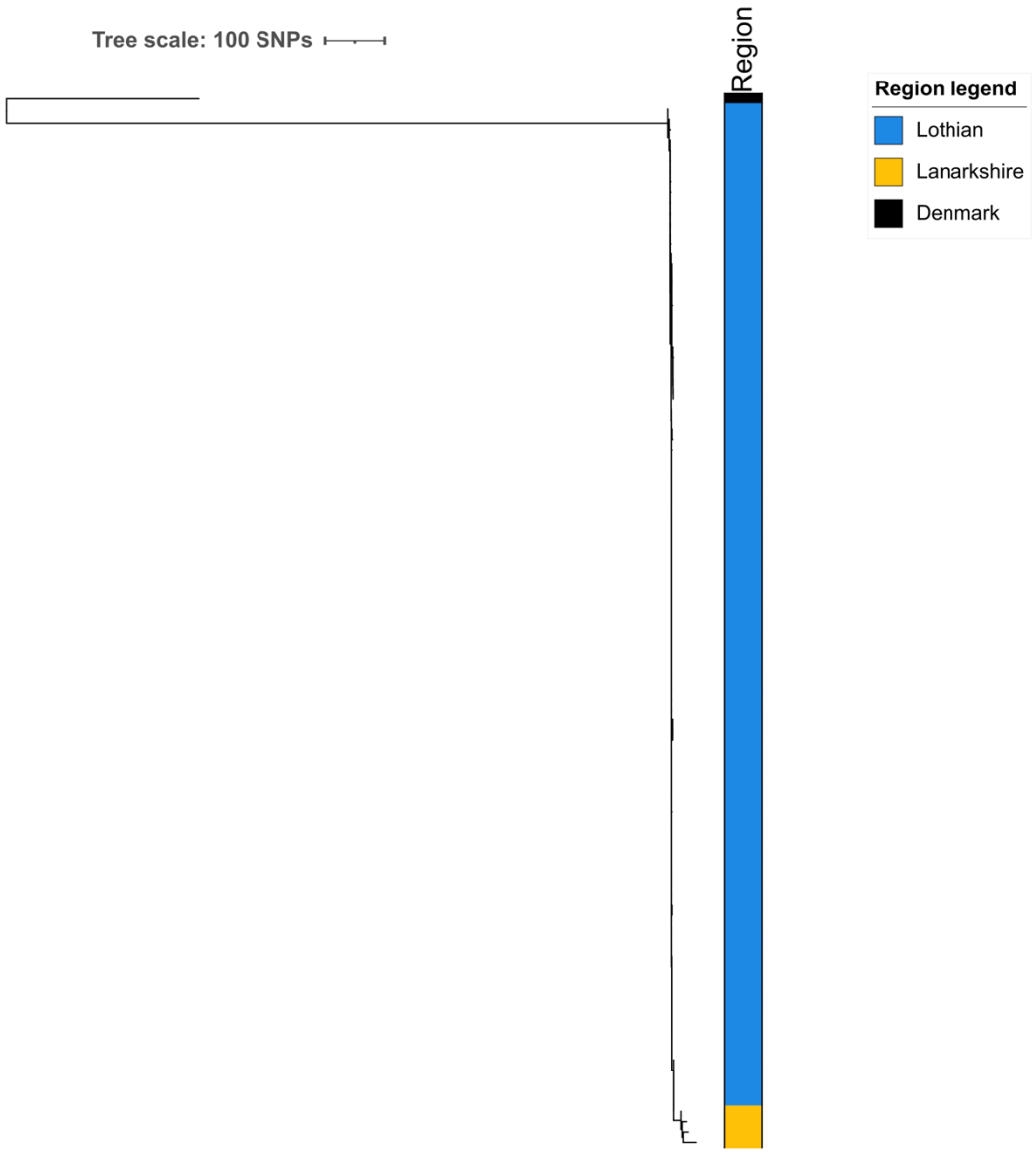


Figure 6.4 Phylogeny of Scottish ST1424 genomes

Genomes (n=102) mapped to V24 chromosome (2,720,495 bp), MGEs and recombination masked (1,427,975 bp), and maximum likelihood phylogeny built. Metadata are indicated

by coloured blocks (see figure for key). Thesis collection: VRE Diversity, Chapter 4; Cluster, Chapter 5; National, Chapter 6.

6.2.5 AMR detection, comparison of genotypic and phenotypic

Prior to inferring AMR patterns in the entire national collection, accuracy of the *in silico* approach was determined by comparing to phenotypic susceptibility results. A subset of 80 isolates from the national collection and the 87 isolates included in Chapter 5 were used, as results from routine phenotypic AST were available.

The FDA provide guidance for the evaluation of AST methods which suggest new tests should have categorical agreement to the Gold Standard method of >89.9%, a ME rate (reference test sensitive, new test resistant) <3%, and a VME rate (reference test resistant, new test sensitive) <2.94% [303]. Considering phenotypic AST as the Gold Standard to compare against genotypic assignment, genotype was acceptable for 5/11 tested antimicrobials including ampicillin, vancomycin, and linezolid (Table 6.6). Teicoplanin was unsuitable due to a high ME rate after detection of *vanA* in seven isolates that tested teicoplanin sensitive, retesting with a different AST method confirmed that six of these isolates were teicoplanin resistant as predicted by the genotype. Chloramphenicol resistance was uncommon based on AST and genotypic detection would be deemed unsuitable for clinical use based on one VME, despite very high categorical agreement. Kanamycin, gentamicin, and tetracycline all had unacceptable major and VME rates. Genotypic determination of trimethoprim susceptibility would also not pass the FDA recommendations due to poor categorical agreement (59.9%) with AST.

Table 6.6 Accuracy of *in silico* antimicrobial susceptibility determination (n=167)

Drug	**Phenotypic AST Resistance, n (%)**	**Genotypic Resistance, n (%)**	**Categorical Agreement, % (95 % CI)**	**Major error, n (%)**	**Very major error, n (%)**
Ampicillin	166 (99.4)	165 (98.8)	99.4 (98.2-100)	0 (0)	1 (0.6)
Vancomycin	166 (99.4)	165 (98.8)	99.4 (98.2-100)	0 (0)	1 (0.6)
Erythromycin	167 (100)	166 (99.4)	99.4 (98.2-100)	0 (0)	1 (0.6)
Chloramphenicol	15 (9)	14 (8.4)	99.4 (98.2-100)	0 (0)	1 (6.7)
Linezolid	0 (0)	1 (0.6)	98.8 (97.1-100)	1 (0.6)	0 (0)
Streptomycin	126 (75.4)	124 (74.3)	98.8 (97.1-100)	0 (0)	2 (1.6)
Teicoplanin	160 (95.8)	165 (98.8)	94.6 (91.2-98.0)	7 (100)	2 (1.3)
Kanamycin	159 (95.2)	153 (91.6)	92.8 (88.9-96.7)	3 (37.5)	9 (5.7)

Tetracycline	154 (92.2)	137 (82)	88.6 (83.8-93.4)	1 (7.7)	18 (11.7)
Gentamicin	123 (73.7)	99 (59.3)	62.9 (55.6-70.2)	19 (43.2)	43 (35.0)
Trimethoprim	90 (53.9)	41 (24.6)	59.9 (52.5-67.3)	NA[a]	NA[a]

[a] No sensitive class for trimethoprim, only intermediate or resistant. Lack of AMR gene was considered intermediate, detection of a gene was considered resistant. There were 67 (40.1%) minor errors where the different call was between intermediate and resistant. No other antibiotics had intermediate reference test results so minor errors were not calculated.

AST, antimicrobial sensitivity test; CI, confidence interval; n, number; NA, not available

6.2.6 *In silico* AMR detection in national collection

While Section 6.2.5 highlights some limitations in defining phenotypic susceptibility from genotypic markers and may preclude the introduction of the method into clinical use for the full AST panel at the current time, the detection of AMR markers from WGS data is still informative to characterise the national collection. *In silico* detection of AMR markers showed the Scottish isolates to carry a median of 10 (range 2-15) AMR markers (Table 6.7). *vanA* was the most common vancomycin resistance mechanism, detected in 96.6% of

isolates while *vanB* was only detected in nine (2.8%) cases and eight of these also carried *vanA*. No vancomycin resistance mechanism was detected in 10 isolates, eight were available for repeat MIC testing which confirmed five were vancomycin sensitive (MIC ≤0.75 mg/l, resistant breakpoint >4 mg/l) and three were confirmed as resistant (MIC >256 mg/l). *De novo* assemblies were made of the three discrepant isolates and uploaded to the CARD and ResFinder AMR detection sites, again no complete vancomycin resistance operons were detected.

Two isolates lacked the *pbp5* ampicillin resistance mutations but were confirmed phenotypically resistant (MIC >256 mg/l, resistance breakpoint >8 mg/l), investigation of *de novo* assemblies could not identify an intact copy of *pbp5*. Isolates lacking *pbp5* have been infrequently detected, but these are usually ampicillin sensitive [49]. Linezolid resistance mutations in the 23S rRNA gene were detected in three isolates. There are six copies of the 23S rRNA gene in *E. faecium*, ~80% reads matched the mutant for VRE_ABD_036 suggesting five mutated copies, and ~30% matched the mutant for VRE_ABD_038 and VRE_EDI_084 suggesting two mutated copies. A 23S rRNA mutant ratio of 80% correlates with clinical resistance, but a ratio of 30% may be sensitive or resistant by phenotypic AST [73,418]. Fluoroquinolone resistance mutations were detected in 99.4% *gyrA* and *parC* sequences, only one isolate had no mutations in either gene. The most common *gyrA* mutation was S83I (n=228, 70.4%), then S83Y (n=94, 29.0%), with S83R and E87G detected in single isolates (0.3%). In *parC* the S80I (n=180, 55.6%) and S80R (n=144, 44.4%) mutations were detected. These mutations all lead to resistance to clinically active

concentrations of fluoroquinolones [419]. The ResFinder database also included genes associated with disinfectant tolerance, these compounds are increasingly used in healthcare and other settings to reduce microbial load on patients and the environment. Quaternary ammonium compound (QAC) resistance genes are multidrug efflux pumps and are also activate against chlorhexidine gluconate. Of these QAC resistance genes only *qacZ* was detected in a single isolate.

Table 6.7 Presence of AMR Markers in Scottish isolates (n=326)

Antimicrobial Class	**Drug**	**Genetic Marker**	**Count (%)**
Beta-lactams	Ampicillin	*pbp5*-R	324 (99.4)
Glycopeptides	Vancomycin, Teicoplanin	*vanA*	315 (96.6)
	Vancomycin	*vanB*	9 (2.8)
Oxazolidinones	Linezolid	23S rRNA G2576T	3 (0.9)
Aminoglycosides	Gentamicin*	*aac(6')-Ii*	326 (100)
	Gentamicin, Kanamycin	*aac (6')-Ie-aph(2'')-Ia*	159 (48.8)
	Kanamycin	*aph (3')-IIIa* or *aac (6')-Ie- aph (2'')*	284 (87.1)
	Streptomycin	*ant(6)-Ia*	244 (74.8)
	Spectinomycin	*ant(9)-Ia*	8 (2.5)
Fluoroquinolones	Ciprofloxacin	*gyrA* mutations	324 (99.4)
	Ciprofloxacin	*parC* mutations	324 (99.4)
Macrolides	Erythromycin	*msrC*	322 (98.8)

	Erythromycin	*ermA*	8 (2.5)
	Erythromycin	*ermB*	290 (89)
	Erythromycin	*ermT*	44 (13.5)
Tetracyclines	Tetracycline	*tet*(M)	171 (52.5)
	Tetracycline	*tet*(S)	66 (20.2)
	Tetracycline	*tet*(L)	36 (11)
Diaminopyrimidines	Trimethoprim	*dfrG*	93 (28.5)
Phenicols	Chloramphenicol	*cat*	19 (5.8)
Lincosamides	Clindamycin	*lsaE*	17 (5.2)
	Clindamycin	*lnuB*	7 (2.1)
Disinfectants	Benzalkonium chloride, Chlorhexidine digluconate	*qacZ*	1 (0.3)

* Confers low level resistance, gentamicin can still be used in higher doses

6.2.7 Plasmid *rep* typing

Plasmid *rep* genes were sought as a marker of plasmid carriage. All isolates showed evidence of plasmid presence, with a median of four (range 2-10) plasmid *rep* genes detected. A total of 16 *rep* genes were detected, with four detected in >75% of isolates (Table 6.8).

Table 6.8 Presence of plasmid *rep* genes (n=326)

***rep* gene**	**Count (%)**
repUS15	321 (98.5)
rep2	304 (93.3)
rep11a	269 (82.5)
rep17	248 (76.1)
rep18b	97 (29.8)
repUS43	78 (23.9)
repUS12	64 (19.6)
repUS7	42 (12.9)
rep1	40 (12.3)
rep14a	27 (8.3)
rep14b	22 (6.7)
rep29	22 (6.7)
repUS57	5 (1.5)
rep7a	1 (0.3)
rep7b	1 (0.3)
repUS56	1 (0.3)

6.2.8 Presence of virulence markers

Virulence markers were screened in the Scottish collection and identified genes encoding the collagen-binding adhesin *acm* in 99.7% (n=325), endocarditis antigen adhesin *efaA* in 92.3% (n=301), hyaluronidase *hyl* in 54.6% (n=178), and surface protein *esp* in 0.3% (n=1).

6.2.9 Association of genetic markers with MLST

Associations between the presence of AMR, plasmid, and virulence markers with MLST were determined using Chi-square (Table 6.9, Figure 6.5). ST203 and ST80 were analysed individually, all other STs were analysed together due to low numbers. Only markers present in ≥20 and ≤306 genomes were analysed to ensure an acceptable sample size of positive and negative cases in the Chi-square analysis.

Table 6.9 Incidence of genetic markers of AMR, plasmids, and virulence in MLST groups

Target				**ST203 (n=125)**	**ST80 (n=117)**	**Other STs (n=84)**
AMR	Beta-lactams	Ampicillin	*pbp5*-R	125 (100)	115 (98.3)	84 (100)
	Glycopeptides	Vancomycin, Teicoplanin	*vanA*	123 (98.4)	114 (97.4)	80 (95.2)

		Vancomycin	*vanB*	0 (0)	1 (0.9)	8 (9.5)
	Oxazolidinones	Linezolid	23S rRNA G2576T	1 (0.8)	2 (1.7)	0 (0)
	Aminoglycosides	Gentamicin	*aac(6')-Ii*	125 (100)	117 (100)	81 (96.4)
		Gentamicin, Kanamycin	*aac(6')-Ie-aph(2'')-Ia**	41 (32.8)	80 (68.4)	38 (45.2)
		Kanamycin	*aph (3')-IIIa* or *aac(6')-Ie-aph(2'')**	121 (96.8)	109 (93.2)	54 (64.3)
		Streptomycin	*ant(6)-Ia**	116 (92.8)	94 (80.3)	34 (40.5)
		Spectinomycin	*ant(9)-Ia*	3 (2.4)	0 (0)	5 (6)
	Macrolides	Erythromycin	*msrC*	125 (100)	116 (99.1)	81 (96.4)
		Erythromycin	*ermA*	3 (2.4)	0 (0)	5 (6)
		Erythromycin	*ermB**	122 (97.6)	106 (90.6)	62 (73.8)
		Erythromycin	*ermT**	0 (0)	19 (16.2)	25 (29.8)
	Lincosamides	Clindamycin	*lnuB*	2 (1.6)	0 (0)	5 (6)
		Clindamycin	*lsaE*	6 (4.8)	1 (0.9)	10 (11.9)
	Tetracyclines	Tetracycline	*tet*(M)*	119 (95.2)	33 (28.2)	19 (22.6)

		Tetracycline	*tet*(S)*	0 (0)	64 (54.7)	2 (2.4)
		Tetracycline	*tet*(L)*	6 (4.8)	0 (0)	30 (35.7)
	Diaminopyrimidines	Trimethoprim	*dfrG**	0 (0)	27 (23.1)	66 (78.6)
	Phenicols	Chloramphenicol	*cat*	1 (0.8)	0 (0)	18 (21.4)
	Disinfectants	Benzalkonium chloride, Chlorhexidine digluconate	*qacZ*	0 (0)	0 (0)	1 (1.2)
	Fluoroquinolones	Ciprofloxacin	*gyrA* S83Y*	1 (0.8)	28 (23.9)	65 (77.4)
		Ciprofloxacin	*gyrA* S83I*	124 (99.2)	89 (76.1)	15 (17.9)
		Ciprofloxacin	*gyrA* S83R	0 (0)	0 (0)	1 (1.2)
		Ciprofloxacin	*gyrA* E87G	0 (0)	0 (0)	1 (1.2)
		Ciprofloxacin	*parC* S80I*	1 (0.8)	116 (99.1)	63 (75)
		Ciprofloxacin	*parC* S80R*	123 (98.4)	1 (0.9)	20 (23.8)
Plasmids			*rep1**	0 (0)	38 (32.5)	2 (2.4)

	*rep2**	125 (100)	113 (96.6)	63 (75)
	rep7a	1 (0.8)	0 (0)	0 (0)
	rep7b	0 (0)	0 (0)	1 (1.2)
	*rep11a**	114 (91.2)	116 (99.1)	39 (46.4)
	*rep14a**	6 (4.8)	4 (3.4)	17 (20.2)
	*rep14b**	3 (2.4)	4 (3.4)	15 (17.9)
	*rep17**	121 (96.8)	83 (70.9)	44 (52.4)
	*rep18b**	18 (14.4)	23 (19.7)	56 (66.7)
	*rep29**	0 (0)	1 (0.9)	21 (25)
	*repUS7**	3 (2.4)	35 (29.9)	4 (4.8)
	*repUS12**	6 (4.8)	20 (17.1)	38 (45.2)
	repUS15	124 (99.2)	114 (97.4)	83 (98.8)
	*repUS43**	0 (0)	42 (35.9)	36 (42.9)
	repUS56	0 (0)	0 (0)	1 (1.2)
	repUS57	0 (0)	0 (0)	5 (6)

Virulence	*acm*	124 (99.2)	117 (100)	84 (100)
	*efaA**	116 (92.8)	109 (93.2)	76 (90.5)
	*hyl**	118 (94.4)	26 (22.2)	34 (40.5)
	esp	1 (0.8)	0 (0)	0 (0)

* Included in chi-square analysis

A significant association ($p<0.001$) was found between the analysed MLST groups and AMR markers, plasmid *rep* genes, and virulence genes. Residuals within the chi-square analysis were investigated to identify which markers were most associated with MLST groups (Figure 6.5, Appendix 3). This showed that *tet* genes were MLST specific with *tet*(M) more common in ST203, *tet*(S) in ST80, and *tet*(L) in other STs. The trimethoprim resistance gene *dfrG* was strongly associated with other STs and underrepresented in ST203. Fluoroquinolone resistance mutations also partitioned with STs, *gyrA* S83Y was enriched in other STs and underrepresented in ST203, while the opposite was true for S83I. In *parC* S80R was strongly associated with ST203 and not ST80, and vice versa for S80I. Associations between plasmid *rep* genes and ST groups were also identified: *rep17* was positively associated with ST203; *rep1* and *repUS7* were positively associated with ST80; *rep14a/b*, *rep18b*, *rep29*, and *repUS12* were positively associated with other STs. The virulence genes had lower strength of association, although *hyl* did have a positive association with ST203 and negative association with ST80.

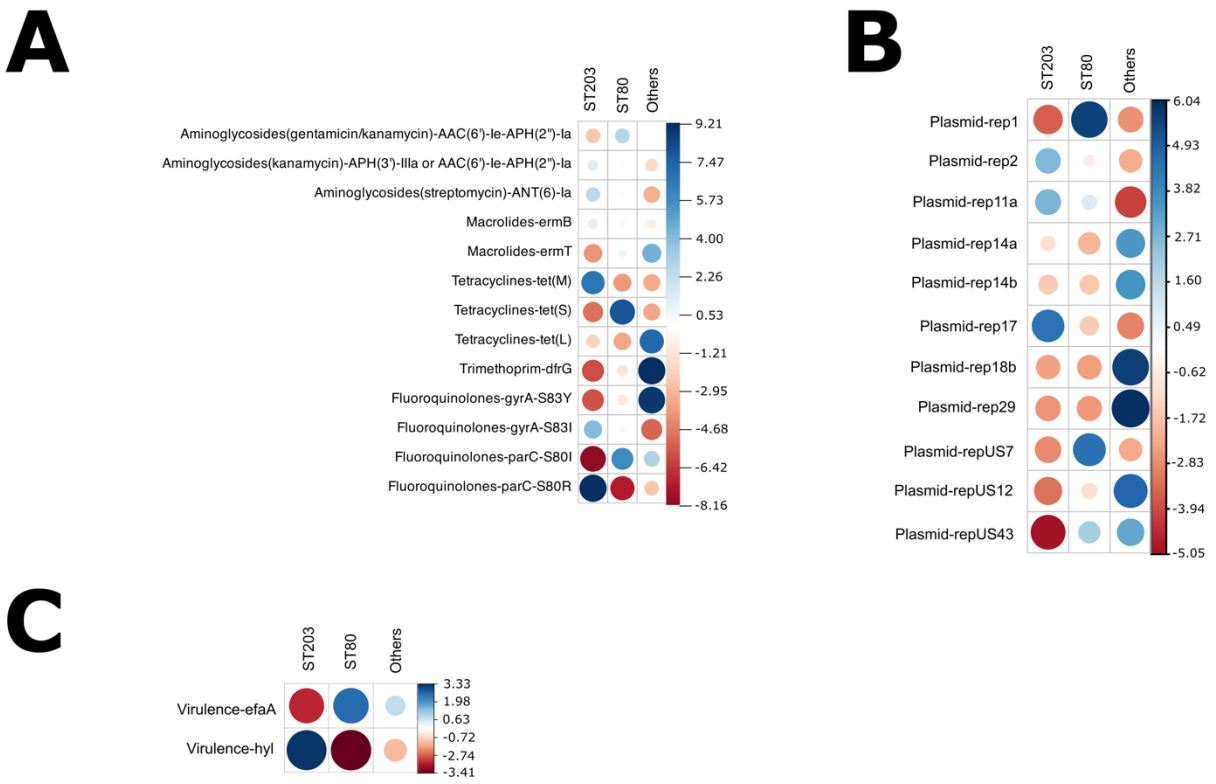


Figure 6.5 Bubble plots of Chi-square residuals

A, AMR markers; B, plasmid *rep* genes; C, virulence genes. Residuals show how far the observed data deviates from the expected number if all categories were equal. The larger the deviation, the larger the residual. Bubble size and colour intensity reflect the size of the residual, negative values showing a negative association (less observations than expected), positive values showing a positive association (more observations than expected).

6.2.10 Genomic comparison between Scottish and international genomes

To further contextualise VREfm in Scotland, 1584 raw short read sequence sets from international studies of *E. faecium* were downloaded from the NCBI [140,283]. The international collection encompassed the years 1946-2016 and 37 countries, 1418 were from the nosocomial Clade A1. Quality trimmed read sets for the 326 Scottish and 1584 international isolates were then used to assign VLKCs with PopPUNK [185]. The 1910 genomes were assigned to 513 VLKCs, 400 (78.0%) were singletons (Appendix 4). Scottish genomes were assigned to 19 VLKCs (median size 2, range 1-146) and 90% of the genomes were found in five VLKCs. In comparison, the global genomes were assigned to 475 VLKCs (median size 1, range 1-755) and 90% of genomes were found in 327 VLKCs. Only eight VLKCs contained genomes from both collections, and most were small with <100 genomes from each collection (Appendix 4). One VLKC (6_12_17_23_30) accounted for 44.8% (n=146) of Scottish and 47.7% (n=755) of international genomes. VLKC 6_12_17_23_30 contained genomes sampled between 1991 and 2016 from 21 countries. The genomes were assigned to 33 MLSTs, ST203 was the most common ST and eight STs accounted for 90% of all genomes (Table 6.10). All 901 genomes assigned to this VLKC were further analysed.

Table 6.10 Sequence Types identified in VLKC 6_12_17_23_30

MLST	International (%) [n=755]	Scottish (%) [n=146]	Total (%) [n=901]	Total Cumulative %
203	175 (23.2)	125 (85.6)	300 (33.3)	33.3
17	124 (16.4)	2 (1.4)	126 (14.0)	47.3
18	99 (13.1)	0 (0)	99 (11.0)	58.3
796	70 (9.3)	0 (0)	70 (7.8)	66.0
78	62 (8.2)	2 (1.4)	64 (7.1)	73.1
192	58 (7.7)	0 (0)	58 (6.4)	79.6
412	45 (6)	5 (3.4)	50 (5.6)	85.1
117	38 (5)	2 (1.4)	40 (4.4)	89.6
341	17 (2.3)	0 (0)	17 (1.9)	91.5
252	15 (2)	0 (0)	15 (1.7)	93.1
400	12 (1.6)	0 (0)	12 (1.3)	94.5
80	10 (1.3)	0 (0)	10 (1.1)	95.6
2228	0 (0)	8 (5.5)	8 (0.9)	96.5
233	6 (0.8)	0 (0)	6 (0.7)	97.1
132	3 (0.4)	0 (0)	3 (0.3)	97.5
414	3 (0.4)	0 (0)	3 (0.3)	97.8
323	2 (0.3)	0 (0)	2 (0.2)	98.0
1005	2 (0.3)	0 (0)	2 (0.2)	98.2
1043	2 (0.3)	0 (0)	2 (0.2)	98.4
204	1 (0.1)	0 (0)	1 (0.1)	98.6

549	1 (0.1)	0 (0)	1 (0.1)	98.7
555	1 (0.1)	0 (0)	1 (0.1)	98.8
780	1 (0.1)	0 (0)	1 (0.1)	98.9
991	1 (0.1)	0 (0)	1 (0.1)	99.0
1032	1 (0.1)	0 (0)	1 (0.1)	99.1
1038	1 (0.1)	0 (0)	1 (0.1)	99.2
1039	1 (0.1)	0 (0)	1 (0.1)	99.3
1042	1 (0.1)	0 (0)	1 (0.1)	99.4
1201	1 (0.1)	0 (0)	1 (0.1)	99.5
1483	1 (0.1)	0 (0)	1 (0.1)	99.7
1486	1 (0.1)	0 (0)	1 (0.1)	99.8
2227	0 (0)	1 (0.7)	1 (0.1)	99.9
2230	0 (0)	1 (0.7)	1 (0.1)	100.0

The ST203 complete reference genome (Aus0085, accession NC_021994) was used for mapping the 900 quality-trimmed short read sets. Aus0085 was included in the initial PopPUNK analysis and was part of VLKC 6_12_17_23_30 so the short reads for this isolate were removed to avoid self-mapping. The mapped genomes were aligned and putative MGEs masked with Snippy, and recombination masked with Gubbins. The optimal root was identified in the Gubbins tree and a root-to-tip analysis performed with BactDating [308]. The root-to-tip analysis showed a significant temporal signal within the dataset (R^2=0.52, $p<1.00x10^{-4}$). The temporal signal within VLKC 6_12_17_23_30 was then fully

inferred using the Bayesian framework in BactDating (Figure 6.6). The most recent common ancestor of VLKC 6_12_17_23_30 was estimated in 1956 (95% CI 1948-1964), and the substitution rate was estimated at 5.7 (95% CI 5.3-6.2) substitutions per genome per year. Most of the Scottish genomes (136/146, 93.2%) clustered together with six genomes sampled in the UK, the ancestral node was dated in 2005 (95% CI 2003-2006). The Scottish cluster was differentiated from other genomes by 32 SNPs in three recombination blocks (542783-543423, 8 SNPs; 1894193-1893929 7 SNPs; 1972840-1973292 17 SNPs) and a further five SNPs outside of recombination blocks (Table 6.11).

Table 6.11 Defining SNPs for the Scottish Cluster in VLKC 6_12_17_23_30

SNP	**Effect**	**CDS**	**CDS Product**
G660242A	Nonsynonymous	EFAU085_RS03060	ABC transporter substrate-binding protein
G1262352A	-	None	-
A1287787C	Nonsynonymous	EFAU085_RS06380	DNA internalization-related competence protein ComEC/Rec2
C2304701A	Nonsynonymous	EFAU085_RS11475	ABC transporter permease
G2346860T	Synonymous	EFAU085_RS11660	ABC transporter ATP-binding protein

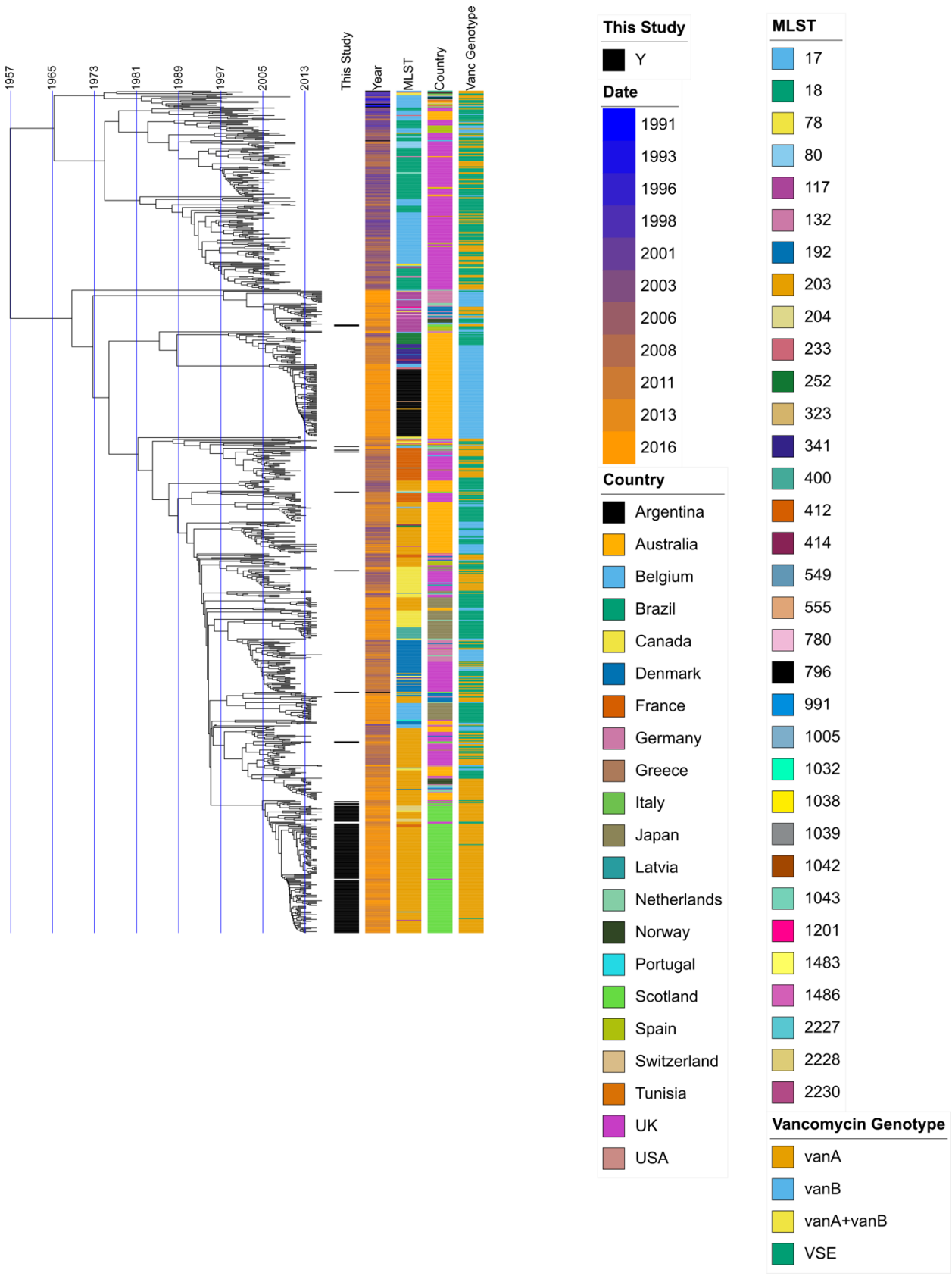

1957
1965
1973
1981
1989
1997
2005
2013
This Study
Year
MLST
Country
Vanc Genotype
This Study
Y
Date
1991
1993
1996
1998
2001
2003
2006
2008
2011
2013
2016
Country
Argentina
Australia
Belgium
Brazil
Canada
Denmark
France
Germany
Greece
Italy
Japan
Latvia
Netherlands
Norway
Portugal
Scotland
Spain
Switzerland
Tunisia
UK
USA
MLST
17
18
78
80
117
132
192
203
204
233
252
323
341
400
412
414
549
555
780
796
991
1005
1032
1038
1039
1042
1043
1201
1483
1486
2227
2228
2230
Vancomycin Genotype
vanA
vanB
vanA+vanB
VSE

Figure 6.6 Dated phylogeny of VLKC 6_12_17_23_30

Genomes (n=900) were mapped to the Aus0085 chromosome (2,994,661 bp), MGEs and recombination masked (588,350 bp), and maximum likelihood phylogeny built. Branch lengths correspond to phylogenetic dating with the internal scale indicated. Metadata are indicated in the tree based on coloured blocks (see figure for key). VSE, vancomycin-sensitive *Enterococcus.*

Six putative alcohol tolerance markers have previously been identified in successful *E. faecium* lineages, including ST203 [310]. We sought to identify alcohol tolerance markers in the genome collection and generated a composite score based on whether markers were detected or not, and the inferred effect of each marker (increased or decreased tolerance, Table 2.1). To avoid confounding from differences in AMR patterns, only *vanA* positive genomes were included in this analysis. Within VLKC 6_12_17_23_30, the 136 genomes in the Scottish clade (134 from this study and two from the international collection) had a mean $\pm$ standard deviation (SD) alcohol tolerance score of 2.1 $\pm$ 0.5, which was not significantly different to the other *vanA*-positive genomes in VLKC 6_12_17_23_30 (n=139, mean $\pm$ SD alcohol tolerance 2.1 $\pm$ 0.7). Within the whole Scottish collection, mean $\pm$ SD alcohol tolerance score for the 134 *vanA*-positive genomes in VLKC 6_12_17_23_30 was 2.1 $\pm$ 0.5, compared to 0.9 $\pm$ 1.0 for the 181 *vanA*-positive Scottish genomes not in VLKC 6_12_17_23_30 ($p < 0.001$).

6.3 Discussion

This chapter describes the genetic diversity of 326 VREfm from Scotland collected during 2012-17, at this time vancomycin resistance in bloodstream isolates increased by 15% so it is of public health importance to understand the drivers of this increase. ST203 and ST80 accounted for 74% of Scottish VREfm, longitudinal studies in Denmark, Germany, and Spain have identified ST203 being the dominant lineage during 2000-2009 before being displaced in the following 10 years by ST117 and ST80 [257,285,420,421]. Scottish ST203 were

grouped with 755 international genomes by PopPUNK, although maximum likelihood phylogeny showed the Scottish genomes were closely related to only six other genomes collected in the UK with the cluster differentiated by 5 genomic SNPs (four in CDS involved in membrane transport or DNA internalisation, Table 6.11) and 32 SNPs in recombination blocks. Indeed, dating analysis suggested a common ancestor of Scottish ST203 in 2005 (95% CI 2003-2006) suggesting the lineage had been present in Scotland for around a decade (Figure 6.6). Around the early 2000s and 2010s significant changes in IPC were occurring in Scottish healthcare due to the impact of MRSA and *C. difficile* - including promoting hand hygiene, use of alcohol hand rub, antimicrobial stewardship, reduction in co-amoxiclav/3rd generation cephalosporins/fluoroquinolone use, and standardisation of practice nationally [422–424]. Putative alcohol tolerance markers were enriched in ST203 and VLKC 6_12_17_23_30 as a whole, but were much less common in non-ST203 Scottish VREfm. Australian ST203 isolates have also been shown to carry these markers which may point to an intrinsic characteristic of the lineage that aids survival in healthcare settings [310]. The identified alcohol tolerance markers confer the ability to survive up to 23% isopropanol *in vitro*, isolates are still killed by 70% isopropanol so alcohol hand rubs when used correctly are still effective for decontamination [310,425].

Scottish ST80 genomes were split between five VLKCs: 1_6_21 contained 94 Scottish genomes (6 SNP clusters 5, 6, 8, 9, 10, and outliers, Figure 6.3) and 14 international (Australia, Germany, Netherlands, and UK), VLKC 8 contained 20 Scottish genomes only (6

SNP cluster 11 and outliers). These VLKCs were not analysed further due to there being <100 genomes from each collection.

Using a threshold of ≤6 core SNPs clusters of related genomes could be identified within Scottish VREfm that highlighted transmission within and between different regions (Figure 6.3, Table 6.5). Cluster 17 was the largest (n=45) and represented ongoing transmission within Tayside, with one genome from Fife suggesting spread into the neighbouring Health Board. Tayside hosts a large teaching hospital which provides specialist services for residents of Fife so patient transfers between the two boards are common [426]. Cluster 9 contained 36 genomes, mainly from Fife but also cases from four other Health Boards showing a wide dispersal across Scotland. Cluster 10 from Dumfries and Galloway (DG) also appears to have arisen from within Cluster 9, Fife and DG rarely transfer patients directly so this may highlight the role of inter-board transfers for specialist care in the national transmission of VREfm [426], or cases linking these two clusters that have been missed in the sampling for this study. Previous studies have highlighted the role of patient transfers within healthcare networks in the regional and national transmission of *E. faecium* in Australia, England, and Germany [140,411,427].

Analysis of the outbreak genomes from Chapters 4 and 5 identified these were mostly related to existing lineages in Lothian, further highlighting the important role of intra-regional transmission patterns driving the burden of VREfm (Figure 6.3). Additionally, there were some instances of phylogenetic clustering between outbreak genomes from

Chapter 4 and 5 within ST80, highlighting the spread of closely related strains with outbreak potential between two hospitals and patient groups (haematology and orthopaedics) within Lothian. On the other hand, ST1424 was only identified in 6 patients from the collection presented in Chapter 4 and was not closely related to contemporaneous ST1424 isolates from Lanarkshire (Figure 6.4).

In healthcare settings, bacterial WGS is largely used as a typing tool to support IPC investigations and identify circulating clones. WGS also provides the capacity to identify genetic AMR markers and so could be used to inform clinical decision making for treatment, if sequencing results can be returned fast enough [428]. In this Chapter, the accuracy of using Ariba and the ResFinder database to infer antimicrobial susceptibility was compared to currently used phenotypic tests for *E. faecium*. Ampicillin, vancomycin, and linezolid are important for the treatment of *E. faecium* and *in silico* susceptibility detection showed high accuracy (Table 6.6). Gentamicin susceptibility determination was poor in this analysis which would preclude *in silico* guidance of VREfm endocarditis treatment, but otherwise aminoglycosides would not usually be considered for *E. faecium* treatment due to the high ampicillin resistance rates ruling out synergy between beta-lactams and aminoglycosides. Teicoplanin did not have acceptable performance due to seven false positive *in silico* calls, repeat phenotypic testing of these isolates showed six isolates were in fact resistant and so were falsely negative on Vitek. Teicoplanin is not commonly used as directed treatment against *E. faecium*, particularly those with vancomycin resistance so the clinical significance of this finding may not be significant.

Trimethoprim accuracy was particularly low, although this antibiotic is not recommended for treatment of enterococcal infections due to the ability to absorb environmental folate leading to poor correlation between *in vitro* MIC and clinical outcomes [37]. Daptomycin and ciprofloxacin are not included on the Vitek card used for phenotypic testing so were not considered in this analysis. The ResFinder database was published in 2013 and included relevant enterococcal resistance genes [302], and WGS has been successfully used to infer susceptibility in other pathogens, in particular *Mycobacterium tuberculosis* where this is now replacing phenotypic AST in reference laboratories [429]. However, only in the past five years have studies began to emerge comparing *in silico* and phenotypic AMR determination in enterococci [190,430–433]. ResFinder is commonly used, and studies are generally small in size (100-200 genomes, low numbers of some phenotypes). A recent preprint from Coll *et al* [433] moves the field forward by presenting a curated database of enterococcal AMR determinants against 15 antimicrobials and an Ariba-based pipeline to detect markers in WGS datasets. The method was evaluated with 4730 *E. faecium* with WGS and phenotypic antimicrobial susceptibility results and outperformed the AMR-Finder, CARD, and ResFinder databases. In comparison to phenotypic results the curated database showed high accuracy for ampicillin, vancomycin, ciprofloxacin, and linezolid. Like the results presented in this Chapter accuracy for teicoplanin, aminoglycosides, and tetracycline was reduced due to MEs, although often due to errors in original phenotypic test or silent/inactivate AMR genes. The curated database included AMR markers for daptomycin and tigecycline, but sensitivity was poor (<40%) highlighting that the mechanisms of resistance against these last-line agents are still to be fully understood.

Despite the challenges highlighted here, Sherry *et al* [434] recently published an ISO:15189 accredited workflow to detect important AMR determinants in a range of pathogenic bacteria (including enterococci), classify the determinants into antibiotic classes, and provide customised reports. Sherry *et al* [434] show that *in silico* AMR detection can be implemented in a clinical setting and this will be an area of active development in the coming years.

Associations were identified between MLST and plasmid *rep* genes, suggesting some plasmids were delineated within STs in Scotland (Table 6.9, Figure 6.5). This may also explain why some AMR genes were enriched in specific STs, if they are carried on particular plasmids. This observation is supported by results in Chapter 4 where limited evidence of plasmid sharing between ST1424 and ST80 was identified. However, a large study of full plasmid sequences by Arredondo-Alonso *et al* [225] identified shared plasmid content between two isolates is highly dependent on the source of the isolates and the collection time, geographic distance had little impact on plasmid relatedness, highlighting that within a transmission network (e.g. healthcare) plasmid contents are highly mobile and spread rapidly. However, the same authors show that enterococcal plasmids are highly modular and can harbour multiple *rep* genes, so the detections in this Chapter may not all map to single plasmids. To fully investigate plasmid contents, long read sequencing is required to overcome repetitive elements (Chapters 3 and 4) [435].

There are some limitations to this work. The Scottish VREfm isolates are not a fully representative, structured sample of the entire population, but represent a convenience sample of available isolates. Efforts were made to avoid oversampling specific Health Boards at the expense of others but the risk of bias due to differences in sample collection dates, and clinical sample types remains. For this reason, the results presented here should not be considered the complete picture of VREfm in Scotland during the sampling period. Nevertheless, there is no clear clustering with regards to time or sample type (Figure 6.2) suggesting the sample is heterogenous and represents some of the diversity within the population. This work only focusses on VREfm, although it is clear from Figure 6.6 VSEfm are dispersed throughout the phylogeny and interspersed with VREfm indicating vancomycin resistance is a fluid marker and by focussing on VREfm some important links may be missed [249]. Currently, IPC is largely focussed on AMR phenotypes and so VREfm are screened for and efforts made to control their transmission, the inclusion of VSEfm in future surveillance would allow monitoring of emerging lineages that may acquire vancomycin resistance in the future, and delineate more transmission networks that could be controlled before they spread widely [393].

To conclude, this Chapter shows that VREfm in Scotland are polyclonal, but dominated by a few STs that have spread within and between Health Boards. Common STs in Scotland are recognised as successful nosocomial lineages around the World, although clustering with international genomes found few close relationships. Further analysis of the predominant Scottish ST203 suggested the lineage was introduced into Scotland around

2005, at which time IPC practice was undergoing profound change. It is likely the prevalent VREfm lineages were well placed to survive in the hospital environment, WGS can play an important role as part of wide-ranging Public Health surveillance to monitor the bacterial populations and detect new threats.

Chapter 7 General discussion and future work

7.1 General Discussion

*E. faecalis a*nd *E. faecium* are leading causes of nosocomial infections, mainly due to their capacity to resist antimicrobial treatments and colonise the gut of hospitalised patients. In Scotland, vancomycin resistance in *E. faecium* has been increasing in recent years putting more strain on the remaining antibiotics, mainly linezolid and daptomycin. WGS allows the delineation of possible bacterial transmission pathways, and an in-depth understanding of the causative AMR mechanisms. The central aim of this thesis was to use WGS to understand the molecular epidemiology of antimicrobial resistant enterococci from human healthcare settings in Scotland. As an NHS Clinical Scientist working in clinical microbiology, this PhD has given me the opportunity to investigate the neglected problem in Scotland with VRE and other significant AMR in enterococci. The results presented here suggest that while linezolid resistance in *E. faecalis* is present in diverse genetic backgrounds, nosocomial VREfm were represented by more closely related lineages within Clade A. Closer inspection of VREfm hospital outbreaks showed evidence of on-ward transmission and possible transmission between wards in the same hospital. Analysis of a national VREfm isolate collection supported the role of intra-regional spread but also highlighted some links between Health Boards and possible national transmission pathways.

In Chapter 3, short and long read sequencing were combined to generate near-complete genome assemblies of six *optrA*-positive *E. faecalis*, *optrA* is an emerging resistance mechanism against the last-line antimicrobial linezolid and these isolates were among the first such cases identified in Scotland. The six isolates were assigned to unrelated STs and differed by a median 18,806 SNPs (range 13,909 – 22,272), ruling out a clonal outbreak. *optrA* was identified on unrelated plasmids and there was limited evidence of a shared *optrA* cassette in the identified plasmids, suggesting multiple transmission networks of the *optrA* gene (Figure 3.2). Comparison to international genomes did show relationships to *optrA* cassettes identified in Europe and East Asia from humans, animals, and pet food (Figure 3.3). Although based on only six isolates, these findings suggest multiple seedings from a diverse *optrA* reservoir.

In Chapter 4, within-patient diversity of VREfm carriage populations were investigated. Within-patient diversity can complicate transmission investigations based on single colony picks, so this chapter aimed to identify how diverse the VREfm population is in rectal carriage and determine the optimal number of colonies to use to effectively detect transmission. Analysis of 229 colony picks from 11 patients over a 1-month period identified carriage of up to three population variants in 27% of patients (Table 4.4). Within individual patients, when VREfm genomes belonged to the same ST a maximum of three core genome SNPs were identified. Three transmission clusters were identified involving 10 patients and transmission resolution was reduced when using less than 14 colony picks per screening sample (Figure 4.6 and Table 4.6). Carriage of multiple *E. faecium* lineages

has been identified in other studies investigating carriage in up to five patients, and a large study of 109 haematology inpatients found 58% carried 2-4 *E. faecium* subtypes [267,282,370,376]. The results in Chapter 4 are in keeping with these estimates and show that consideration of within patient diversity impacts on transmission resolution. The use of 14 colonies per sample may not be feasible for routine use as this increases costs and complexity of the final analysis, while only revealing diversity in 27% of patients. Power calculation suggested five colonies would reveal 50% of the within-patient diversity and may be a pragmatic choice for routinely identifying within-patient diversity to reduce costs, and this strategy has been successfully applied elsewhere (Table 4.1) [376]. The degree of within-patient diversity considered in transmission analyses will ultimately be determined by the aim of the IPC investigation - if a high-confidence transmission network is required then 14 (or more) colonies may be necessary, if the presence only of the main outbreak strain(s) in the studied patients is required then fewer colonies can be used, Table 4.1 could be used as a guide for such decisions in the future.

Chapter 4 also highlighted the presence of plasmids and AMR genes was variable between genomes that differed by zero core genome SNPs (Figures 4.2, 4.3, and 4.7). This is perhaps not surprising, given the masking approach taken to generate a core genome removes some of the most variable regions. However, it is a reminder that core SNPs are just one measure of distinguishing genomes. Variable presence of AMR genes was also identified in another studies of within-patient diversity, including the vancomycin resistance gene *vanA* [282]. Chapter 4 also showed close relationships between colonising

populations and subsequent bloodstream isolates, this has also been identified by other studies which did not identify any genetic markers that predispose BSI [12,282,376]. These findings suggest the progression of *E. faecium* from carriage to bloodstream invasion is predicated more by host factors than the gain of a particular marker, and that the combination of patient immune status and the pre-existing predilection of nosocomial *E. faecium* for survival in healthcare settings is sufficient to allow bloodstream invasion [436].

In Chapter 5, WGS was applied to a larger outbreak of VREfm occurring over four months with 84 patients identified as VREfm positive. The aim of this chapter was to investigate the utility of merged WGS and epidemiological analysis to understand suspected VREfm outbreaks. Two WGS analysis approaches were used: core genome SNPs using a suspected transmission cut-off of 3 SNPs (based on maximum SNP distance between isolates of the same lineage within individual patients identified in Chapter 4) and PopPIPE which clusters genomes based on whole genome relationships using SKA. Both WGS methods had higher resolution than PFGE, this finding is in agreement with other studies and confirms WGS is a superior method for outbreak investigations [254,263,264]. Within the WGS methods, core SNPs clustered 85% of genomes into 14 clusters while PopPIPE clustered 89% of genomes into 20 clusters, clustering agreed between the two methods in 78% of cases. Epidemiological support (patients on same ward at same time) was marginally higher for PopPIPE clusters (Table 5.3). The results in this Chapter suggest PopPIPE is a modest improvement on core SNPs, recent studies by Higgs *et al*[305] and Maechler *et al*[401] show a more pronounced improvement in clustering and epidemiological support when using SKA

compared to core SNPs. The differences in strength of findings may be due to the higher core SNP threshold applied in these studies (7 and 10 SNPs, compared to 3 SNPs here) which will cluster more genomes regardless of epidemiological linkage than that applied in Chapter 5, and the application of these approaches to more genomes (308 and 693, compare to 87 here) allowing differences between the methods to be more obvious due to higher statistical power. Together, the data in Chapter 5 and the literature support the use of SKA-based approaches for investigating *E. faecium* outbreaks [305,401]. Chapter 5 highlights the applicability of PopPIPE for this task, as SKA analysis is recommended on genomes that are already known to be closely related so PopPUNK could be used to generate clusters of related genomes prior to sub clustering with PopPIPE. A reference-free *k*-mer based methodology is attractive as it may be quicker to implement into routine use as the validation of reference choice and core genome masking strategy would not be required, and would likely have lower turnaround times in suspected outbreaks as *k*-mer approaches are generally very quick [184,187]. The reference-free PopPIPE approach could also be applied to different bacterial species of interest, providing a unified protocol for WGS-based typing and outbreak investigations.

From the epidemiological data and using timing of VREfm positivity in Chapter 5, 25 patients could be identified as introductions carrying VREfm on admission and three patients were confirmed acquisitions during admission. The acquisitions were in PopPIPE clusters with introductions and on the same ward at the same time, supporting a transmission link. PopPIPE also identified likely transmission clusters in patients that could

not be classified as introductions or acquisitions based on sampling data alone (Figure 5.3). Cluster analysis also showed linked cases in unconnected wards, and patients admitted to the hospital already carrying VREfm belonging to a known cluster with no recognised healthcare contact in the preceding three months (Figure 5.3). These results suggest multiple introductions of VREfm into the hospital, with evidence of transmission within and possibly between wards. Complex pictures of *E. faecium* transmission involving multiple hospital wards have been uncovered in other WGS based studies, this may be due to uncovered links between wards such as movement of staff or equipment or the movement of patients to central facilities such as imaging suites [249,357,407], alternatively this could be due to ongoing introductions driven by healthcare contact or community-based transmission. Current evidence suggests VREfm are not common in healthy individuals in the community, and nosocomial *E. faecium* lineages are infrequently identified in community, animal, or food samples which would point more towards healthcare contact as a driver for transmission [100,102,103,140,437].

In Chapter 6, WGS was applied to a collection of 326 VREfm from eight Scottish regional Health Boards. The aim of this chapter was to identify the genetic background of VREfm across Scotland. MLST identified ST203 and ST80 as the main STs (75% of all genomes), with another 18 STs detected in lower numbers (Table 6.4). For more in depth understanding, the core SNP cut-off applied for direct transmission in Chapter 5 was doubled to 6 SNPs to detect more distantly related cases across different regions. In total, 73% of genomes were assigned to 17 clusters, five clusters (23% of genomes) contained

genomes from a single Health Board (Figure 6.3 and Table 6.9). The remaining 12 clusters involved up to five different Health Boards, indicating widespread dissemination of VREfm lineages. Dispersal of VREfm within referral networks has been described in other countries [140,411,427], patient movements for specialist care within Scotland could explain the clustering of VREfm genomes from different Health Boards.

Given the observation that VREfm transmission between regions in Scotland was not uncommon, investigation as to whether there was sharing of international clones was performed using PopPUNK and a collection of 1584 *E. faecium* genomes from the UK and international studies. A total of 513 VLKCs were present, but Scottish genomes were only present in 19 of these, VLKC 6_12_17_23_30 was the largest containing 146 Scottish and 755 international genomes and was analysed further. A phylogeny based on mapping to a reference genome within VLKC 6_12_17_23_30 showed 134 of the Scottish genomes clustered together with six from the UK, the other 12 Scottish genomes were not closely related to others in the phylogeny (Figure 6.12). The Scottish cluster was predicted to have a common ancestor in 2005 (95% CI 2003-2006) suggesting this lineage has been present in Scotland for at least a decade. ST203 was identified in other European countries around this time and was later displaced by other STs [257,285,420,421]. Putative markers of alcohol tolerance were enriched in VLKC 6_12_17_23_30, which may provide a selective advantage in the healthcare environment [310].

The findings in this thesis shed some light on the lineages driving VREfm in Scotland in the mid 2010s, and hint at some of the factors that made them successful. This dataset is also useful as a backdrop to understand newly identified clusters within Scotland. For example, many of the outbreak genomes identified in Chapters 4 and 5 clustered within pre-existing lineages in Lothian (Figure 6.3). ST1424 was not identified in the national collection but was found in Lothian and Lanarkshire in 2017 (six and four patients respectively, see Chapter 4 and Lemonidis *et al*[284]), while the Lothian cases represented a clonal outbreak the Lanarkshire cases were more diverse likely representing a large population of ST1424 in that Health Board. Our analysis is not powered to detect changes in population variants over time, but it is worth noting the ST1424 lineage did displace successful lineages in other countries around the late 2010s, for example being the most common *E. faecium* lineage in Australian BSIs in 2019 [438]. These observations highlight the potential benefits from ongoing WGS-based surveillance for *E. faecium* in Scotland, and at the time of writing a proposal is being considered by the Public Health Scotland Pathogen Genomics Oversight Group to develop this activity in part informed by the data demonstrated in this thesis.

Chapter 3 investigates some of the first identified *optrA*-positive linezolid resistant *E. faecalis* in Scotland, detected in the community and healthcare. Linezolid resistance due to *optrA* and *poxtA* has been identified in *E. faecalis* from animals, food, and humans with a suggestion that the resistance mechanism is selected by the use of florfenicol in agriculture and transmitted to humans via animal contact or the food chain [353,439,440]. This

has hallmarks of VREfm in the 1990s, which was selected in livestock by the use of the glycopeptide avoparcin for growth promotion and transmitted to humans via the food chain [441,442]. Avoparcin was banned as a growth promoter in livestock in 1997 [443] and a large study from the UK by Gouliouris *et al* found VREfm carriage in livestock fell from 66% in 2003 to 0% in 2014-15 [437,444]. In the same study, the majority of sampled *E. faecium* from livestock animals and human BSIs were not phylogenetically related suggesting transmission between these two sources was uncommon. However, analysis of Clade A1 nosocomial *E. faecium* by van Hal *et al* [252] identified that Clade A2 (animal-associated) and Clade B (commensals) are important reservoirs for adaptation within A1 mediated by recombination driven by MGEs, this highlights that although rare, animal associated *E. faecium* do transfer to humans and can have impacts on nosocomial lineages. *E. faecalis* is more of a generalist and does not display the genetic clades linked to host species seen in *E. faecium*, therefore lineages that do transfer from animals into humans are more readily able to cause disease. Transferable linezolid resistance has been identified in a wide range of Gram-positive bacteria, including VREfm, but was not present in any VREfm in this thesis where linezolid resistant was very rare [357,358].

Chapters 3 and 4 used long read sequencing to generate hybrid genome assemblies. This approach allowed the in-depth investigation of plasmid structures to be performed, which was essential for confirming that *optrA*-positive *E. faecalis* carried different MGEs as opposed to a shared linezolid resistance element and the detection of a linear *vanA* plasmid in *E. faecium*. Hybrid assemblies also allowed the use of phylogenetically close

reference genomes for mapping, this ensures more of the genetic variation is considered in core genome SNP typing which is an important consideration in transmission studies. In comparison, use of public reference genomes that may have been isolated many years prior to the isolates of interest and in different geographic regions may lead to informative genome regions being omitted from the core genome, reducing the ability to resolve close relationships. Long read sequencing is increasingly applied in microbial genomics to identify the genetic context of AMR, investigate MGEs, and produce novel reference genomes [225,258,445,446]. Hybrid assembly is still the optimal approach, although with improvements in long read accuracy, particularly from ONT platforms, the generation of accurate long-read only assemblies may soon be possible [153,154,447]. However, a limitation of long read sequencing over short reads at this time is the throughput. Long read platforms can generate up to 12 or 24 bacterial genomes per flowcell, although the larger ONT PromethION platforms may allow higher throughput these are generally only found in large core sequencing labs that have high throughput to keep the cost per genome down. Illumina provides platforms with a range of throughputs but many flowcells produce more sequence data per run than equivalent ONT sequencers, allowing more genomes to be generated per run. Another challenge with long read sequencing is the need for high amounts of input DNA (generally 200 ng to 1 µg DNA per sample for non-PCR amplified ONT library preparations, compared to ≤100 ng for Illumina) of longer fragment size which requires optimised extraction methods that can be expensive, time consuming, and not easily automated. Another challenge with ONT sequencing is the pace of improvements, with chemistry and bioinformatic methods updating regularly making

implementation into an accredited service challenging. At this time, long read sequencing is best utilised as an adjunct to short read sequencing in laboratories interested in genomic surveillance, with short reads used for bulk sequencing and long reads used to generate complete assemblies of genomes of interest.

There are some overarching limitations of this work. All the included isolates were resistant to linezolid (Chapter 3) or vancomycin (Chapters 4-6), which means the findings may not be representative of wider populations. This may be important if resistance is frequently gained by sensitive strains as suggested by some genomic epidemiology studies [249,282,448,449]. However, recent analysis from a single hospital in Denmark found no correlation between VSEfm and VREfm clones over a four year period suggesting vancomycin resistance acquisition by VSEfm is not always frequent [450]. All the analyses presented in this thesis were retrospective and so were not used to influence patient care. The results of this work provide a valuable insight into enterococcal AMR in Scotland and may inform future IPC planning but to show the full utility of WGS for outbreak management WGS must be performed and fed back to clinical teams as near to real time as possible [407,451].

7.2 Future Work

The data generated in this thesis provides a baseline for future genomic surveillance of AMR in enterococci in Scotland. This activity should become a routine public health

function to inform ongoing IPC policy to reduce the incidence of VREfm. A significant upscale in WGS capacity occurred in response to the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) pandemic and the expertise, equipment, and bioinformatics capacity put in place for this could be applied to bacterial AMR surveillance and other public health threats in future. WGS alone will not impact enterococcal AMR and must be linked to work to improve patient outcomes, antimicrobial utilisation, and infection control measures to reduce the burden of these bacteria.

Enterococcal AMR surveillance should also be extended into animal health, given the possible link between agricultural antibiotic use and the generation of *optrA*-mediated linezolid resistance. A possible means of population-level surveillance of AMR bacteria transfer into humans from agriculture would be to screen cases of *Salmonella*, *Campylobacter*, *Cryptosporidium*, or Shiga toxin positive *E. coli*, as these zoonotic pathogens are often linked to improper food preparation or contact with animals which could also be a means of transfer for enterococci. Another surveillance measure would be to screen wastewater for AMR determinants [452].

As described in Chapter 5, it can be challenging to interpret enterococcal outbreaks with sampling not always coinciding with epidemiological definitions of healthcare acquired infections, and a reduced sensitivity of rectal screening methods. This would be much improved with an agreed guideline on how to approach enterococcal outbreaks giving evidence based advice on sampling strategies. For example, how to sample for

environmental reservoirs, when and how often to screen patients in an outbreak setting, optimal choice of screening approach (solid agar or enrichment broth), and how to approach within-patient diversity. Such a guideline could include recommendations for multiple nosocomial pathogens to increase utility in healthcare settings and provide optimal impact of WGS surveillance.

Future studies should aim to improve the reliability of *in silico* AMR detection in enterococci. A recent preprint shows promise but the approach was only applied against *E. faecium* and showed low accuracy against the last line agents daptomycin and tigecycline [433]. Collaboration between healthcare laboratories and academic centres to identify the underlying resistance mechanisms is essential.

Within patient diversity was analysed in Chapter 4. The choice of 14 colonies was based on statistical power calculation but may not be feasible for routine use due to excessive costs. Future studies applying mixed strain inputs for real time outbreak detection should be investigated, which will allow the diversity present within a sample to be investigated using a single index in a WGS library preparation. Plate sweep metagenomics with the mSWEEP pipeline determines the lineages present and estimates their abundance in a sweep of growth from an agar plate [377], or strain-resolved metagenomics can be attempted directly on clinical samples with the StrainXpress pipeline [453]. These approaches are yet to be applied to genomes from cases of transmission but may form the basis of future developments in bacterial epidemiology.

7.3 Conclusion

In summary, this thesis shows that VREfm in Scotland is driven by multiple clones transmitted within healthcare settings and migrating between regional Health Boards. Two approaches to investigating local outbreaks were presented, one taking account of within-patient diversity to resolve patient linkage with high precision, and another showing the utility of reference-free whole genome analysis to cluster suspected outbreak cases. In comparison to VREfm, linezolid resistant *E. faecalis* due to *optrA* were less common and were identified in patients based in the community and healthcare, suggesting a different means of selection and acquisition. It is likely that different public health strategies will be required to tackle these different threats. This work also highlights the utility of long read sequencing to investigate exceptional AMR patterns and to generate outbreak-specific reference genomes for short read mapping. WGS should be an important part of the public health response against AMR in enterococci.

References

1. Martin JD, Mundt JO. Enterococci in insects. *Appl Microbiol* 1972; **24**: 575–80.

2. Mundt JO. Occurrence of enterococci in animals in a wild environment. *Appl Microbiol* 1963; **11**: 136–40.

3. MacCallum WG, Hastings TW. A case of acute endocarditis caused by micrococcus zymogenes (nov. Spec.), with a description of the microorganism. *J Exp Med* 1899; **4**: 521–34.

4. Thiercelin ME. Morphologie et modes de reproduction de l'enterocoque. *Comptes Rendus Séances Société Biol Ses Fil* 1899; **11**: 551–3.

5. Thiercelin ME, Jouhaud L. Sur un diplococque saprophyte de l'intestin susceptible de devenir pathogene. *Comptes Rendus Séances Société Biol Ses Fil* 1899: 269–71.

6. Schleifer KH, Kilpper-Balz R. Transfer of Streptococcus faecalis and Streptococcus faecium to the Genus Enterococcus nom. rev. as Enterococcus faecalis comb. nov. and Enterococcus faecium comb. nov. *Int J Syst Bacteriol* 1984; **34**: 31–4.

7. Ranotkar S, Kumar P, Zutshi S, *et al.* Vancomycin-resistant enterococci: Troublemaker of the 21st century. *Integr Med Res* 2014; **2**: 205–12.

8. UKHSA. *Laboratory surveillance of Enterococcus spp. bacteraemia (England): 2021*. London, UK: UK Health Security Agency; 2023. Available at:

https://www.gov.uk/government/publications/enterococcus-spp-bacteraemia-voluntary-surveillance-2021/laboratory-surveillance-of-enterococcus-spp-bacteraemia-england-2021.

9. Jiang H-L, Zhou Z, Wang L-S, *et al.* The Risk Factors, Costs, and Survival Analysis of Invasive VRE Infections at a Medical Center in Eastern Taiwan. *Int J Infect Dis* 2016; **54**: 18–24.

10. McKinnell JA, Kunz DF, Moser SA, *et al.* Patient-level analysis of incident vancomycin-resistant enterococci colonization and antibiotic days of therapy. *Epidemiol Infect* 2016; **144**: 1748–55.

11. Billington EO, Phang SH, Gregson DB, *et al.* Incidence, Risk Factors, and Outcomes for Enterococcus spp . Blood Stream Infections: A Population-Based Study. *Int J Infect Dis* 2014; **26**: 76–82.

12. Raven KE, Gouliouris T, Parkhill J, *et al.* Genome-Based Analysis of Enterococcus faecium Bacteremia Associated with Recurrent and Mixed-Strain Infection Carroll KC, ed. *J Clin Microbiol* 2018; **56**: e01520-17.

13. Sievert DM, Ricks P, Edwards JR, *et al.* Antimicrobial-resistant pathogens associated with healthcare-associated infections: summary of data reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2009-2010. *Infect Control Hosp Epidemiol* 2013; **34**: 1–14.

14. Verway M, Brown KA, Marchand-Austin A, *et al.* Prevalence and Mortality Associated with Bloodstream Organisms: a Population-Wide Retrospective Cohort Study. *J Clin Microbiol* 2022; **60**: e02429-21.

15. Brinkwirth S, Ayobami O, Eckmanns T, *et al.* Hospital-acquired infections caused by enterococci: a systematic review and meta-analysis, WHO European Region, 1 January 2010 to 4 February 2020. *Eurosurveillance* 2021; **26**: 2001628.

16. Ikuta KS, Swetschinski LR, Aguilar GR, *et al.* Global mortality associated with 33 bacterial pathogens in 2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet* 2022; **400**: 2221–48.

17. Ong DSY, Bonten MJM, Safdari K, *et al.* Epidemiology, management, and risk-adjusted mortality of ICU-acquired enterococcal bacteremia. *Clin Infect Dis* 2015; **61**: 1413–20.

18. Dik JWH, Dinkelacker AG, Vemer P, *et al.* Cost-analysis of seven nosocomial outbreaks in an academic hospital. *PLoS ONE* 2016; **11**: 1–7.

19. Jung E, Byun S, Lee H, *et al.* Vancomycin-resistant Enterococcus colonization in the intensive care unit: Clinical outcomes and attributable costs of hospitalization. *Am J Infect Control* 2014; **42**: 1062–6.

20. Lloyd-Smith P, Younger J, Lloyd-Smith E, *et al.* Economic analysis of vancomycin-resistant enterococci at a Canadian hospital: Assessing attributable cost and length of stay. *J Hosp Infect* 2013; **85**: 54–9.

21. ARHAI Scotland. *Scottish One Health Antimicrobial Use and Antimicrobial Resistance Report 2021*. Glasgow, UK: ARHAI Scotland; 2022. Available at: https://www.nss.nhs.scot/media/3394/sonaar-2021-report.pdf.

22. García-Solache M, Rice LB. The Enterococcus: A model of adaptability to its environment. *Clin Microbiol Rev* 2019; **32**: e00058-18.

23. Gao W, Howden BP, Stinear TP. Evolution of virulence in Enterococcus faecium, a hospital-adapted opportunistic pathogen. *Curr Opin Microbiol* 2018; **41**: 76–82.

24. Singh KV, Nallapareddy SR, Sillanpää J, *et al.* Importance of the Collagen Adhesin Ace in Pathogenesis and Protection against Enterococcus faecalis Experimental Endocarditis. *PLOS Pathog* 2010; **6**: e1000716.

25. Nallapareddy SR, Singh KV, Okhuysen PC, *et al.* A Functional Collagen Adhesin Gene, acm, in Clinical Isolates of Enterococcus faecium Correlates with the Recent Success of This Emerging Nosocomial Pathogen. *Infect Immun* 2008; **76**: 4110–9.

26. Sillanpää J, Nallapareddy SR, Singh KV, *et al.* Characterization of the ebpfm pilus-encoding operon of Enterococcus faecium and its role in biofilm formation and virulence in a murine model of urinary tract infection. *Virulence* 2010; **1**: 236–46.

27. Chow JW, Thal LA, Perri MB, *et al.* Plasmid-associated hemolysin and aggregation substance production contribute to virulence in experimental enterococcal endocarditis. *Antimicrob Agents Chemother* 1993; **37**: 2474–7.

28. Shankar N, Lockatell CV, Baghdayan AS, *et al.* Role of Enterococcus faecalis surface protein Esp in the pathogenesis of ascending urinary tract infection. *Infect Immun* 2001; **69**: 4366–72.

29. Jett BD, Huycke MM, Gilmore MS. Virulence of Enterococci. *Clin Microbiol Rev* 1994; **7**: 462–78.

30. Park SY, Kim KM, Lee JH, *et al.* Extracellular Gelatinase of Enterococcus faecalis Destroys a Defense System in Insect Hemolymph and Human Serum. *Infect Immun* 2007; **75**: 1861–9.

31. Paganelli FL, Huebner J, Singh KV, *et al.* Genome-wide Screening Identifies Phosphotransferase System Permease BepA to Be Involved in Enterococcus faecium Endocarditis and Biofilm Formation. *J Infect Dis* 2016; **214**: 189–95.

32. Kristich CJ, Rice LB. Enterococcal Infection — Treatment and Antibiotic Resistance. In: Gilmore MS, Clewell DB, Ike Y, et al., eds. *Enterococci: From Commensals to Leading Causes of Drug Resistant Infection [Internet]*. Boston: Massachusetts Eye and Ear Infirmary, 2014. Available at: https://www.ncbi.nlm.nih.gov/books/NBK190420/.

33. Arias CA, Murray BE. The rise of the Enterococcus: beyond vancomycin resistance. *Nat Rev Microbiol* 2012; **10**: 266–78.

34. Hollenbeck BL, Rice LB. Intrinsic and acquired resistance mechanisms in enterococcus. *Virulence* 2012; **3**: 421–33.

35. Cairns KA, Udy AA, Peel TN, *et al.* Therapeutics for Vancomycin-Resistant Enterococcal Bloodstream Infections. *Clin Microbiol Rev* 2023: e00059-22.

36. Murray BE. The life and times of the Enterococcus. *Clin Microbiol Rev* 1990; **3**: 46–65.

37. Wisell KT, Kahlmeter G, Giske CG. Trimethoprim and enterococci in urinary tract infections : new perspectives on an old issue. *J Antimicrob Chemother* 2008; **62**: 35–40.

38. Leavis HL, Willems RJL, Top J, *et al.* High-level ciprofloxacin resistance from point mutations in gyrA and parC confined to global hospital-adapted clonal lineage CC17 of Enterococcus faecium. *J Clin Microbiol* 2006; **44**: 1059–64.

39. Brisse S, Fluit AC, Wagner U, *et al.* Association of alterations in ParC and GyrA proteins with resistance of clinical isolates of Enterococcus faecium to nine different fluoroquinolones. *Antimicrob Agents Chemother* 1999; **43**: 2513–6.

40. Portillo A, Ruiz-Larrea F, Zarazaga M, *et al.* Macrolide resistance genes in Enterococcus spp. *Antimicrob Agents Chemother* 2000; **44**: 967–71.

41. Zalipour M, Esfahani BN, Havaei SA. Phenotypic and genotypic characterization of glycopeptide, aminoglycoside and macrolide resistance among clinical isolates of Enterococcus faecalis: a multicenter based study. *BMC Res Notes* 2019; **12**: 292.

42. Singh KV, Malathum K, Murray BE. 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* 2001; **45**: 263–6.

43. Singh KV, Weinstock GM, Murray BE. An Enterococcus faecalis ABC Homologue (Lsa) Is Required for the Resistance of This Species to Clindamycin and Quinupristin-Dalfopristin. *Antimicrob Agents Chemother* 2002; **46**: 1845–50.

44. O'Driscoll T, Crank CW. Vancomycin-resistant enterococcal infections : epidemiology , clinical manifestations , and optimal management. *Infect Drug Resist* 2015; **8**: 217–30.

45. Rice LB, Desbonnet C, Tait-Kamradt A, *et al.* Structural and Regulatory Changes in PBP4 Trigger Decreased β-Lactam Susceptibility in Enterococcus faecalis. *mBio* 2018; **9**: e00361-18.

46. Pietta E, Montealegre MC, Roh JH, *et al.* Enterococcus faecium PBP5-S/R, the Missing Link between PBP5-S and PBP5-R. *Antimicrob Agents Chemother* 2014; **58**: 6978–81.

47. Rice LB, Carias LL, Rudin S, *et al.* Enterococcus faecium low-affinity pbp5 is a transferable determinant. *Antimicrob Agents Chemother* 2005; **49**: 5007–12.

48. García-Solache M, Lebreton F, McLaughlin RE, *et al.* Homologous recombination within large chromosomal regions facilitates acquisition of beta-lactam and vancomycin resistance in Enterococcus faecium . *Antimicrob Agents Chemother* 2016; **60**: 5777–86.

49. Novais C, Tedim AP, Lanza VF, *et al.* Co-diversification of Enterococcus faecium Core Genomes and PBP5: Evidences of pbp5 Horizontal Transfer. *Front Microbiol* 2016; **7**: 1–17.

50. Rice LB, Marshall SH. Evidence of incorporation of the chromosomal beta-lactamase gene of Enterococcus faecalis CH19 into a transposon derived from staphylococci. *Antimicrob Agents Chemother* 1992; **36**: 1843–6.

51. Habib G, Lancellotti P, Antunes MJ, *et al.* 2015 ESC Guidelines for the management of infective endocarditis: The Task Force for the Management of Infective Endocarditis of the European Society of Cardiology (ESC)Endorsed by: European Association for Cardio-Thoracic Surgery (EACTS), the European Association of Nuclear Medicine (EANM). *Eur Heart J* 2015; **36**: 3075–128.

52. Ramirez MS, Tolmasky ME. Aminoglycoside Modifying Enzymes. *Drug Resist Updat* 2010; **13**: 151–71.

53. Shete V, Grover N, Kumar M. Analysis of Aminoglycoside Modifying Enzyme Genes Responsible for High-Level Aminoglycoside Resistance among Enterococcal Isolates. *J Pathog* 2017; **2017**: 3256952.

54. Health Protection Scotland. *Scottish One Health Antimicrobial Use and Resistance in 2018 Annual Report*. Glasgow, UK: HPS; 2019. Available at: https://www.hps.scot.nhs.uk/web-resources-container/scottish-one-health-antimicrobial-use-and-antimicrobial-resistance-in-2018/.

55. Arthur M, Depardieu F, Gerbaud G, *et al.* The VanS sensor negatively controls VanR-mediated transcriptional activation of glycopeptide resistance genes of Tn1546 and related elements in the absence of induction. *J Bacteriol* 1997; **179**: 97–106.

56. Arthur M, Depardieu F, Molinas C, *et al.* The vanZ gene of Tn1546 from enterococcus faecium BM4147 confers resistance to teicoplanin. *Gene* 1995; **154**: 87–92.

57. Arthur M, Molinas C, Courvalin P. Sequence of the vanY gene required for production of a vancomycin-inducible D,D-carboxypeptidase in Enterococcus faecium BM4147. *Gene* 1992; **120**: 111–4.

58. Reynolds PE, Depardieu F, Dutka-Malen S, *et al.* Glycopeptide resistance mediated by enterococcal transposon Tn 1546 requires production of VanX for hydrolysis of D-alanyl-D-alanine. *Mol Microbiol* 1994; **13**: 1065–70.

59. Holmes NE, Ballard SA, Lam MMC, *et al.* Genomic analysis of teicoplanin resistance emerging during treatment of vanB vancomycin-resistant Enterococcus faecium infections in solid organ transplant recipients including donor-derived cases. *J Antimicrob Chemother* 2013; **68**: 2134–9.

60. Hayden MK, Trenholme GM, Schultz JE, *et al.* In Vivo Development of Teicoplanin Resistance in a VanB Enterococcus faecium Isolate. *J Infect Dis* 1993; **167**: 1224–7.

61. Woodford N, Johnson AP, Morrison D, *et al.* Current perspectives on glycopeptide resistance. *Clin Microbiol Rev* 1995; **8**: 585–615.

62. ECDC. *Antimicrobial resistance in the EU/EEA (EARS-Net) - Annual epidemiological report 2021*. Stockholm, Sweden: European Centre for Disease Prevention and Control; 2022. Available at: https://www.ecdc.europa.eu/en/publications-data/antimicrobial-resistance-eueea-ears-net-annual-epidemiological-report-2020.

63. Carvalhaes CG, Sader HS, Streit JM, *et al.* Activity of Oritavancin against Gram-Positive Pathogens Causing Bloodstream Infections in the United States over 10 Years: Focus on Drug-Resistant Enterococcal Subsets (2010–2019). *Antimicrob Agents Chemother* 2022; **66**: e01667-21.

64. Sacramento AG, Zanella RC, Esposito F, *et al.* Changed epidemiology during intra and interhospital spread of high-risk clones of vanA-containing Enterococcus in Brazilian hospitals. *Diagn Microbiol Infect Dis* 2017; **88**: 348–51.

65. Coombs GW, Daley DA, Yee NWT, *et al.* Australian Group on Antimicrobial Resistance (AGAR) Australian Enterococcal Sepsis Outcome Programme (AESOP) Annual Report 2020. *Commun Dis Intell* 2022; **46**. Available at: https://www1.health.gov.au/internet/main/publishing.nsf/Content/2A15CD097063EF40CA2587CE008354F1/$File/australian_group_on_antimicrobial_resistance_agar_australian_enterococcal_sepsis_outcome_programme_aesop_annual_report_2020.pdf. Accessed May 5, 2022.

66. Sivaradjy M, Gunalan A, Priyadarshi K, *et al.* Increasing Trend of Vancomycin-resistant Enterococci Bacteremia in a Tertiary Care Hospital of South India: A Three-year Prospective Study. *Indian J Crit Care Med* 2021; **25**: 881–5.

67. Alemayehu T, Hailemariam M. Prevalence of vancomycin-resistant enterococcus in Africa in one health approach: a systematic review and meta-analysis. *Sci Rep* 2020; **10**: 20542.

68. Hu F, Zhu D, Wang F, *et al.* Current Status and Trends of Antibacterial Resistance in China. *Clin Infect Dis* 2018; **67**: S128–34.

69. Health Protection Scotland and Information Services Devision. *Scottish Antimicrobial Use and Resistance in Humans in 2015*. Glasgow, UK: HPS and ISD; 2016. Available at: https://www.isdscotland.org/Health-Topics/Prescribing-and-Medicines/Publications/2016-08-30/2016-08-30-SAPG-2015-Report.pdf.

70. Moellering RC. Linezolid: The First Oxazolidinone Antimicrobial. *Ann Intern Med* 2003; **138**: 135–42.

71. Burdette SD, Trotman R. Tedizolid: The First Once-Daily Oxazolidinone Class Antibiotic Saravolatz LD, ed. *Clin Infect Dis* 2015; **61**: 1315–21.

72. Hasman H, Clausen PTLC, Kaya H, *et al.* LRE-Finder, a Web tool for detection of the 23S rRNA mutations and the optrA, cfr, cfr(B) and poxtA genes encoding linezolid resistance in enterococci from whole-genome sequences. *J Antimicrob Chemother* 2019; **74**: 1473–6.

73. Marshall SH, Donskey CJ, Hutton-Thomas R, *et al.* Gene Dosage and Linezolid Resistance in Enterococcus faecium and Enterococcus faecalis. *Antimicrob Agents Chemother* 2002; **46**: 3334–6.

74. Pillai SK, Sakoulas G, Wennersten C, *et al.* Linezolid Resistance in *Staphylococcus aureus:* Characterization and Stability of Resistant Phenotype. *J Infect Dis* 2002; **186**: 1603–7.

75. Deshpande LM, Ashcraft DS, Kahn HP, *et al.* Detection of a new *cfr*-like gene, *cfr*(B), in *Enterococcus faecium* isolates recovered from human specimens in the United States as part of the SENTRY Antimicrobial Surveillance Program. *Antimicrob Agents Chemother* 2015; **59**: 6256–61.

76. Diaz L, Kiratisin P, Mendes RE, *et al.* Transferable plasmid-mediated resistance to linezolid due to *cfr* in a human clinical isolate of *Enterococcus faecalis*. *Antimicrob Agents Chemother* 2012; **56**: 3917–22.

77. Guerin F, Sassi M, Dejoies L, *et al.* Molecular and functional analysis of the novel cfr(D) linezolid resistance gene identified in Enterococcus faecium. *J Antimicrob Chemother* 2020; **75**: 1699–703.

78. Wang Y, Lv Y, Cai J, *et al.* A novel gene, *optrA*, that confers transferable resistance to oxazolidinones and phenicols and its presence in *Enterococcus faecalis* and *Enterococcus faecium* of human and animal origin. *J Antimicrob Chemother* 2015; **70**: 2182–90.

79. Antonelli A, D'Andrea MM, Brenciani A, *et al.* Characterization of *poxtA*, a novel phenicol–oxazolidinone–tetracycline resistance gene from an MRSA of clinical origin. *J Antimicrob Chemother* 2018; **73**: 1763–9.

80. Deshpande LM, Castanheira M, Flamm RK, *et al.* Evolving oxazolidinone resistance mechanisms in a worldwide collection of enterococcal clinical isolates: Results from the SENTRY Antimicrobial Surveillance Program. *J Antimicrob Chemother* 2018; **73**: 2314–22.

81. Health Protection Scotland. Oxazolidinone-resistance due to optrA in Enterococcus faecalis. *HPS Wkly Rep* 2016; **50**: 230–1.

82. Arias CA, Panesso D, McGrath DM, *et al.* Genetic Basis for In Vivo Daptomycin Resistance in Enterococci. *N Engl J Med* 2011; **365**: 892–900.

83. Diaz L, Tran TT, Munita JM, *et al.* Whole-Genome Analyses of Enterococcus faecium Isolates with Diverse Daptomycin MICs. *Antimicrob Agents Chemother* 2014; **58**: 4527–34.

84. Miller WR, Bayer AS, Arias CA. Mechanism of Action and Resistance to Daptomycin in Staphylococcus aureus and Enterococci. *Cold Spring Harb Perspect Med* 2016; **6**: a026997.

85. Humphries RM, Pollett S, Sakoulas G. A current perspective on daptomycin for the clinical microbiologist. *Clin Microbiol Rev* 2013; **26**: 759–80.

86. Humphries RM. The new, new daptomycin breakpoint for *Enterococcus* spp. *J Clin Microbiol* 2019; **57**: e00600-19.

87. Britt NS, Potter EM, Patel N, *et al.* Comparison of the Effectiveness and Safety of Linezolid and Daptomycin in Vancomycin-Resistant Enterococcal Bloodstream Infection: A National Cohort Study of Veterans Affairs Patients. *Clin Infect Dis* 2015; **61**: 871–8.

88. Chuang YC, Lin HY, Yang JL, *et al.* Influence of daptomycin doses on the outcomes of VRE bloodstream infection treated with high-dose daptomycin. *J Antimicrob Chemother* 2022; **77**: 2278–87.

89. Lai C-C, Sheng W-H, Wang J-T, *et al.* Safety and efficacy of high-dose daptomycin as salvage therapy for severe gram-positive bacterial sepsis in hospitalized adult patients. *BMC Infect Dis* 2013; **13**: 66.

90. Speer BS, Shoemaker NB, Salyers AA. Bacterial resistance to tetracycline: mechanisms, transfer, and clinical significance. *Clin Microbiol Rev* 1992; **5**: 387–99.

91. Wang J, Pan Y, Shen J, *et al.* The efficacy and safety of tigecycline for the treatment of bloodstream infections: a systematic review and meta-analysis. *Ann Clin Microbiol Antimicrob* 2017; **16**: 24.

92. Fiedler S, Bender JK, Klare I, *et al.* Tigecycline resistance in clinical isolates of *Enterococcus faecium* is mediated by an upregulation of plasmid-encoded tetracycline determinants *tet* (L) and *tet* (M). *J Antimicrob Chemother* 2016; **71**: 871–81.

93. Humphreys H. Controlling the spread of vancomycin-resistant enterococci. Is active screening worthwhile? *J Hosp Infect* 2014; **88**: 191–8.

94. Alatorre-Fernández P, Mayoral-Terán C, Velázquez-Acosta C, *et al.* A polyclonal outbreak of bloodstream infections by Enterococcus faecium in patients with hematologic malignancies. *Am J Infect Control* 2017; **45**: 260–6.

95. Caballero S, Kim S, Carter RA, *et al.* Cooperating Commensals Restore Colonization Resistance to Vancomycin-Resistant Enterococcus faecium Article Cooperating Commensals Restore Colonization Resistance to Vancomycin-Resistant Enterococcus faecium. *Cell Host Microbe* 2017; **21**: 592–602.

96. Iosifidis E, Evdoridou I, Agakidou E, *et al.* Vancomycin-resistant Enterococcus outbreak in a neonatal intensive care unit: Epidemiology, molecular analysis and risk factors. *Am J Infect Control* 2013; **41**: 857–61.

97. Brown AR, Amyes SGB, Paton R, *et al.* Epidemiology and control of vancomycin-resistant enterococci (VRE) in a renal unit. *J Hosp Infect* 1998; **40**: 115–24.

98. Quale J, Landman D, Atwood E, *et al.* Experience with a hospital-wide outbreak of vancomycin-resistant enterococci. *Am J Infect Control* 1996; **24**: 372–9.

99. Alevizakos M, Gaitanidis A, Nasioudis D, *et al.* Colonization With Vancomycin-Resistant Enterococci and Risk for Bloodstream Infection Among Patients With Malignancy: A Systematic Review and Meta-Analysis. *Open Forum Infect Dis* 2017; **4**: ofw246.

100. Whelton E, Lynch C, O'Reilly B, *et al.* Vancomycin-resistant enterococci carriage in an acute Irish hospital. *J Hosp Infect* 2016; **93**: 175–80.

101. Wilson HJ, Khokhar F, Enoch DA, *et al.* Point-prevalence survey of carbapenemase-producing Enterobacteriaceae and vancomycin-resistant enterococci in adult inpatients in a university teaching hospital in the UK. *J Hosp Infect* 2018; **100**: 35–9.

102. Croughan S, O'Cronin D, O'Brien D, *et al.* Vancomycin-Resistant Enterococci in Patients Attending for Colonoscopy: An Estimate of Community Prevalence. *Ir Med J* 2022; **115**: 649.

103. Huang Y-S, Lai L-C, Chen Y-A, *et al.* Colonization With Multidrug-Resistant Organisms Among Healthy Adults in the Community Setting: Prevalence, Risk Factors, and Composition of Gut Microbiome. *Front Microbiol* 2020; **11**: 1402.

104. Noskin GA, Stosor V, Cooper I, *et al.* Recovery of vancomycin-resistant enterococci on fingertips and environmental surfaces. *Infect Control Hosp Epidemiol* 1995; **16**: 577–81.

105. Decker BK, Lau AF, Dekker JP, *et al.* Healthcare personnel intestinal colonization with multidrug-resistant organisms. *Clin Microbiol Infect* 2017; **24**: 82.E1-82.E4.

106. Stellfox ME, Van Tyne D. Last Bacteria Standing: VREfm Persistence in the Hospitalized Gut. *mBio* 2022; **13**: e00670-22.

107. Huang C, Feng S, Huo F, *et al.* Effects of Four Antibiotics on the Diversity of the Intestinal Microbiota. *Microbiol Spectr* 2022; **10**: e01904-21.

108. Ubeda C, Taur Y, Jenq RR, *et al.* Vancomycin-resistant Enterococcus domination of intestinal microbiota is enabled by antibiotic treatment in mice and precedes bloodstream invasion in humans. *J Clin Invest* 2010; **120**: 4332–41.

109. Kitsios GD, Morowitz MJ, Dickson RP, *et al.* Dysbiosis in the ICU: Microbiome science coming to the bedside. *J Crit Care* 2017; **38**: 84–91.

110. Cheng VC, Chen JH, Tai JW, *et al.* Decolonization of gastrointestinal carriage of vancomycin-resistant Enterococcus faecium: case series and review of literature. *BMC Infect Dis* 2014; **14**: 514.

111. De Waele JJ, Leroux-Roels I, Depuydt P. Selective digestive decontamination - Pro. *Intensive Care Med* 2023; **49**: 979–81.

112. Hurley JC. Studies of selective digestive decontamination as a natural experiment to evaluate topical antibiotic prophylaxis and cephalosporin use as population-level risk factors for enterococcal bacteraemia among ICU patients. *J Antimicrob Chemother* 2019; **74**: 3087–94.

113. Wang B, Briegel J, Krueger WA, *et al.* Ecological effects of selective oral decontamination on multidrug-resistance bacteria acquired in the intensive care unit: a case-control study over 5 years. *Intensive Care Med* 2022; **48**: 1165–75.

114. Bilsen MP, Lambregts MMC, van Prehn J, *et al.* Faecal microbiota replacement to eradicate antimicrobial resistant bacteria in the intestinal tract - a systematic review. *Curr Opin Gastroenterol* 2022; **38**: 15–25.

115. Frost I, Sati H, Garcia-Vello P, *et al.* The role of bacterial vaccines in the fight against antimicrobial resistance: an analysis of the preclinical and clinical development pipeline. *Lancet Microbe* 2023; **4**: e113–25.

116. Pradal I, Casado A, Del Rio B, *et al.* Enterococcus faecium Bacteriophage vB_EfaH_163, a New Member of the Herelleviridae Family, Reduces the Mortality Associated with an E. faecium vanR Clinical Isolate in a Galleria mellonella Animal Model. *Viruses* 2023; **15**: 179.

117. Kunz Coyne AJ, Stamper K, El Ghali A, *et al.* Phage-Antibiotic Cocktail Rescues Daptomycin and Phage Susceptibility against Daptomycin-Nonsusceptible Enterococcus faecium in a Simulated Endocardial Vegetation Ex Vivo Model. *Microbiol Spectr* 2023: e0034023.

118. Almeida-Santos AC, Novais C, Peixe L, *et al.* Enterococcus spp. as a Producer and Target of Bacteriocins: A Double-Edged Sword in the Antimicrobial Resistance Crisis Context. *Antibiotics* 2021; **10**: 1215.

119. Karki S, Land G, Aitchison S, *et al.* Long-Term Carriage of Vancomycin-Resistant Enterococci in Patients Discharged from Hospitals: a 12-Year Retrospective Cohort Study. *J Clin Microbiol* 2013; **51**: 3374–9.

120. Sohn KM, Peck KR, Joo EJ, *et al.* Duration of colonization and risk factors for prolonged carriage of vancomycin-resistant enterococci after discharge from the hospital. *Int J Infect Dis* 2013; **17**: e240–6.

121. Wagenvoort JHT, De Brauwer EIGB, Penders RJR, *et al.* Environmental survival of vancomycin-sensitive ampicillin-resistant Enterococcus faecium (AREfm). *Eur J Clin Microbiol Infect Dis* 2015; **34**: 1901–3.

122. Kramer A, Schwebke I, Kampf G. How long do nosocomial pathogens persist on inanimate surfaces? A systematic review. *BMC Infect Dis* 2006; **6**: 130.

123. Michael KE, No D, Roberts MC. vanA-positive multi-drug-resistant Enterococcus spp. isolated from surfaces of a US hospital laundry facility. *J Hosp Infect* 2017; **95**: 218–23.

124. Muzslay M, Moore G, Turton JF, *et al.* Dissemination of antibiotic-resistant enterococci within the ward environment: The role of airborne bacteria and the risk posed by unrecognized carriers. *Am J Infect Control* 2013; **41**: 57–60.

125. Bonten MJM, Hayden MK, Nathan C, *et al.* Epidemiology of colonisation of patients and environment with vancomycin-resistant enterococci. *Lancet* 1996; **348**: 1615–9.

126. Bressan R, Knezevich A, Monticelli J, *et al.* Spread of Vancomycin-Resistant Enterococcus faecium Isolates Despite Validated Infection Control Measures in an Italian Hospital: Antibiotic Resistance and Genotypic Characterization of the Endemic Strain. *Microb Drug Resist* 2018; **24**: 1148–55.

127. Frakking FNJ, Bril WS, Sinnige JC, *et al.* Recommendations for the successful control of a large outbreak of vancomycin-resistant Enterococcus faecium in a non-endemic hospital setting. *J Hosp Infect* 2018; **100**: e216–25.

128. Weber DJ, Kanamori H, Rutala WA. 'No touch' technologies for environmental decontamination: focus on ultraviolet devices and hydrogen peroxide systems. *Curr Opin Infect Dis* 2016; **29**: 424–31.

129. Blane B, Coll F, Raven K, *et al.* Impact of a new hospital with close to 100% single-occupancy rooms on environmental contamination and incidence of vancomycin-resistant Enterococcus faecium colonisation or infection: a genomic surveillance study. *J Hosp Infect* 2023; **139**: 192–200.

130. Fisher D, Pang L, Salmon S, *et al.* A Successful Vancomycin-Resistant Enterococci Reduction Bundle at a Singapore Hospital. *Infect Control Hosp Epidemiol* 2016; **37**: 107–9.

131. Isenman H, Fisher D. Advances in prevention and treatment of vancomycin-resistant Enterococcus infection. *Curr Opin Infect Dis* 2016; **29**: 577–82.

132. Olesen B, Juhl-Jørgensen A, Tronier S, *et al.* A new successful approach to combating vancomycin-resistant enterococcus. *J Hosp Infect* 2015; **91**: 375–6.

133. Lucet JC, Armand-Lefevre L, Laurichesse JJ, *et al.* Rapid control of an outbreak of vancomycin-resistant enterococci in a French university hospital. *J Hosp Infect* 2007; **67**: 42–8.

134. Bearman G, Abbas S, Masroor N, *et al.* Impact of Discontinuing Contact Precautions for Methicillin-Resistant Staphylococcus aureus and Vancomycin-Resistant Enterococcus: An Interrupted Time Series Analysis. *Infect Control Hosp Epidemiol* 2018; **39**: 676–82.

135. Haessler S, Martin EM, Scales ME, *et al.* Stopping the routine use of contact precautions for management of MRSA and VRE at three academic medical centers: An interrupted time series analysis. *Am J Infect Control* 2020; **48**: 1466–73.

136. Morgan DJ, Murthy R, Munoz-Price LS, *et al.* Reconsidering Contact Precautions for Endemic Methicillin-Resistant Staphylococcus aureus and Vancomycin-Resistant Enterococcus. *Infect Control Hosp Epidemiol* 2015; **36**: 1163–72.

137. Ruiz-Garbajosa P, Bonten MJM, Robinson DA, *et al.* Multilocus sequence typing scheme for *Enterococcus faecalis* reveals hospital-adapted genetic complexes in a background of high rates of recombination. *J Clin Microbiol* 2006; **44**: 2220–8.

138. Homan WL, Tribe D, Poznanski S, *et al.* Multilocus Sequence Typing Scheme for Enterococcus faecium. *J Clin Microbiol* 2002; **40**: 1963–71.

139. Carter GP, Buultjens AH, Ballard SA, *et al.* Emergence of endemic MLST non-typeable vancomycin-resistant Enterococcus faecium. *J Antimicrob Chemother* 2016; **71**: 3367–71.

140. Raven KE, Reuter S, Reynolds R, *et al.* A decade of genomic history for healthcare-associated Enterococcus faecium in the United Kingdom and Ireland. *Genome Res* 2016; **26**: 1388–96.

141. Bezdicek M, Hanslikova J, Nykrynova M, *et al.* New Multilocus Sequence Typing Scheme for Enterococcus faecium Based on Whole Genome Sequencing Data. *Microbiol Spectr* 2023; **ACCEPTED**: e05107-22.

142. Turabelidze D, Kotetishvili M, Jr JGM, *et al.* Improved Pulsed-Field Gel Electrophoresis for Typing Vancomycin-Resistant Enterococci. *J Clin Microbiol* 2000; **38**: 4242–5.

143. Titze-de-Almeida R, Willems RJL, Top J, *et al.* Multilocus variable-number tandem-repeat polymorphism among Brazilian Enterococcus faecalis strains. *J Clin Microbiol* 2004; **42**: 4879–81.

144. Top J, Schouls LM, Bonten MJM, *et al.* Multiple-Locus Variable-Number Tandem Repeat Analysis, a Novel Typing Scheme To Study the Genetic Relatedness and Epidemiology of Enterococcus faecium Isolates. *J Clin Microbiol* 2004; **42**: 4503–11.

145. Werner G, Klare I, Witte W. The current MLVA typing scheme for Enterococcus faecium is less discriminatory than MLST and PFGE for epidemic-virulent, hospital-adapted clonal types. *BMC Microbiol* 2007; **7**: 28.

146. Seng P, Drancourt M, Gouriet F, *et al.* Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clin Infect Dis* 2009; **49**: 543–51.

147. Schlebusch S, Price GR, Gallagher RL, *et al.* MALDI-TOF MS meets WGS in a VRE outbreak investigation. *Eur J Clin Microbiol Infect Dis* 2017; **36**: 495–9.

148. Freitas AR, Sousa C, Novais C, *et al.* Rapid detection of high-risk Enterococcus faecium clones by MALDI-TOF mass spectrometry. *Diagn Microbiol Infect Dis* 2016; **2**: 1–9.

149. Griffin PM, Price GR, Schooneveldt JM, *et al.* Use of matrix-assisted laser desorption ionization-time of flight mass spectrometry to identify vancomycin-resistant enterococci and investigate the epidemiology of an outbreak. *J Clin Microbiol* 2012; **50**: 2918–31.

150. Quainoo S, Coolen JPM, van Hijum SAFT, *et al.* Whole-Genome Sequencing of Bacterial Pathogens: the Future of Nosocomial Outbreak Analysis. *Clin Microbiol Rev* 2017; **30**: 1015–63.

151. Bentley DR, Balasubramanian S, Swerdlow HP, *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 2008; **456**: 53–9.

152. Ku C-S, Roukos DH. From next-generation sequencing to nanopore sequencing technology: paving the way to personalized genomic medicine. *Expert Rev Med Devices* 2013; **10**: 1–6.

153. Sereika M, Kirkegaard RH, Karst SM, *et al.* Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. *Nat Methods* 2022; **19**: 823–6.

154. Zhao W, Zeng W, Pang B, *et al.* Oxford nanopore long-read sequencing enables the generation of complete bacterial and plasmid genomes without short-read sequencing. *Front Microbiol* 2023; **14**: 1179966.

155. Kellogg EA. Genome sequencing: Long reads for a short plant. *Nat Plants* 2015; **1**: 15169.

156. Rhoads A, Au KF. PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics* 2015; **13**: 278–89.

157. Cock PJA, Fields CJ, Goto N, *et al.* The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res* 2010; **38**: 1767–71.

158. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013. Available at: http://arxiv.org/abs/1303.3997. Accessed August 16, 2023.

159. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012; **9**: 357–9.

160. Li H. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* 2016; **32**: 2103–10.

161. Li H, Handsaker B, Wysoker A, *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; **25**: 2078–9.

162. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. 2012. Available at: http://arxiv.org/abs/1207.3907. Accessed August 16, 2023.

163. Croucher NJ, Page AJ, Connor TR, *et al.* Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* 2015; **43**: e15.

164. Didelot X, Wilson DJ. ClonalFrameML: Efficient Inference of Recombination in Whole Bacterial Genomes. *PLoS Comput Biol* 2015; **11**: 1–18.

165. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014; **30**: 1312–3.

166. Price MN, Dehal PS, Arkin AP. FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS ONE* 2010; **5**.

167. Minh BQ, Schmidt HA, Chernomor O, *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* 2020; **37**: 1530–4.

168. Ronquist F, Teslenko M, van der Mark P, *et al.* MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Syst Biol* 2012; **61**: 539–42.

169. Zerbino DR, Birney E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 2008; **18**: 821–9.

170. Bankevich A, Nurk S, Antipov D, *et al.* SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol* 2012; **19**: 455–77.

171. Souvorov A, Agarwala R, Lipman DJ. SKESA: strategic k-mer extension for scrupulous assemblies. *Genome Biol* 2018; **19**: 153.

172. Koren S, Walenz BP, Berlin K, *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* 2017; **27**: 722–36.

173. Kolmogorov M, Yuan J, Lin Y, *et al.* Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 2019; **37**: 540–6.

174. Shafin K, Pesout T, Lorig-Roach R, *et al.* Nanopore sequencing and the Shasta toolkit enable efficient de novo assembly of eleven human genomes. *Nat Biotechnol* 2020; **38**: 1044–53.

175. Wick RR, Judd LM, Holt KE. Assembling the perfect bacterial genome using Oxford Nanopore and Illumina sequencing. *PLOS Comput Biol* 2023; **19**: e1010905.

176. Wick RR, Judd LM, Gorrie CL, *et al.* Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 2017; **13**: e1005595.

177. Wick RR, Judd LM, Cerdeira LT, *et al.* Trycycler: consensus long-read assemblies for bacterial genomes. *Genome Biol* 2021; **22**: 266.

178. Watson M, Warr A. Errors in long-read assemblies can critically affect protein prediction. *Nat Biotechnol* 2019; **37**: 124–6.

179. Walker BJ, Abeel T, Shea T, *et al.* Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLOS ONE* 2014; **9**: e112963.

180. Vaser R, Sović I, Nagarajan N, *et al.* Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res* 2017; **27**: 737–46.

181. Seemann T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 2014; **30**: 2068–9.

182. Schwengers O, Jelonek L, Dieckmann MA, *et al.* Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microb Genomics* 2021; **7**: 000685.

183. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol* 2019; **20**: 257.

184. Uelze L, Grützke J, Borowiak M, *et al.* Typing methods based on whole genome sequencing data. *One Health Outlook* 2020; **2**: 3.

185. Lees JA, Harris SR, Tonkin-Hill G, *et al.* Fast and flexible bacterial genomic epidemiology with PopPUNK. *Genome Res* 2019; **29**: 304–16.

186. Gardner SN, Hall BG. When Whole-Genome Alignments Just Won't Work: kSNP v2 Software for Alignment-Free SNP Discovery and Phylogenetics of Hundreds of Microbial Genomes. *PLoS ONE* 2013; **8**: e81760.

187. Harris SR. SKA: Split Kmer Analysis Toolkit for Bacterial Genomic Epidemiology. 2018. Available at: https://www.biorxiv.org/content/10.1101/453142v1. Accessed July 28, 2023.

188. Inouye M, Dashnow H, Raven L-A, *et al.* SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med* 2014; **6**: 1–16.

189. Hunt M, Mather AE, Sánchez-Busó L, *et al.* ARIBA: Rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb Genomics* 2017; **3**: e000131.

190. Bortolaia V, Kaas Rs, Ruppe E, *et al.* ResFinder 4.0 for predictions of phenotypes from genotypes. *J Antimicrob Chemother* 2020; **75**: 3491–500.

191. Alcock BP, Huynh W, Chalil R, *et al.* CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Res* 2023; **51**: D690–9.

192. Feldgarden M, Brover V, Gonzalez-Escalona N, *et al.* AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Sci Rep* 2021; **11**: 12728.

193. Joensen Kg, Scheutz F, Lund O, *et al.* Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli. *J Clin Microbiol* 2014; **52**: 1501–10.

194. Chen L, Zheng D, Liu B, *et al.* VFDB 2016: hierarchical and refined dataset for big data analysis--10 years on. *Nucleic Acids Res* 2016; **44**: D694-697.

195. Carattoli A, Zankari E, Garciá-Fernández A, *et al.* In Silico detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 2014; **58**: 3895–903.

196. van der Graaf-van Bloois L, Wagenaar JA, Zomer AL. RFPlasmid: predicting plasmid sequences from short-read assembly data using machine learning. *Microb Genomics* 2021; **7**: 000683.

197. Arredondo-Alonso S, Rogers MRC, Braat JC, *et al.* mlplasmids: a user-friendly tool to predict plasmid- and chromosome-derived sequences for single species. *Microb Genomics* 2018; **4**: e000224.

198. de Been M, Pinholt M, Top J, *et al.* Core Genome Multilocus Sequence Typing Scheme for High-Resolution Typing of Enterococcus faecium. *J Clin Microbiol* 2015; **53**: 3788–97.

199. Maiden MCJ, Jansen van Rensburg MJ, Bray JE, *et al.* MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat Rev Microbiol* 2013; **11**: 728–36.

200. McInerney JO, McNally A, O'Connell MJ. Why prokaryotes have pangenomes. *Nat Microbiol* 2017; **2**: 17040.

201. Stevenson C, Hall JP, Harrison E, *et al.* Gene mobility promotes the spread of resistance in bacterial populations. *ISME J* 2017; **11**: 1930–2.

202. Mikalsen T, Pedersen T, Willems R, *et al.* Investigating the mobilome in clinically important lineages of Enterococcus faecium and Enterococcus faecalis. *BMC Genomics* 2015; **16**: 282.

203. Kim EB, Marco ML. Nonclinical and clinical Enterococcus faecium strains, but not Enterococcus faecalis strains, have distinct structural and functional genomic features. *Appl Environ Microbiol* 2014; **80**: 154–65.

204. Lam MMC, Seemann T, Bulach DM, *et al.* Comparative analysis of the first complete Enterococcus faecium genome. *J Bacteriol* 2012; **194**: 2334–41.

205. Paulsen IT. Role of Mobile DNA in the Evolution of Vancomycin-Resistant Enterococcus faecalis. *Science* 2003; **299**: 2071–4.

206. Palmer KL, Gilmore MS. Multidrug-Resistant Enterococci Lack CRISPR-cas. *mBio* 2010; **1**: e00227-10.

207. Price VJ, Huo W, Sharifi A, *et al.* CRISPR-Cas and Restriction-Modification Act Additively against Conjugative Antibiotic Resistance Plasmid Transfer in Enterococcus faecalis. *mSphere* 2016; **1**: e00064-16.

208. Gilmore MS, Lebreton F, van Schaik W. Genomic transition of enterococci from gut commensals to leading causes of multidrug-resistant hospital infection in the antibiotic era. *Curr Opin Microbiol* 2013; **16**: 10–6.

209. Lebreton F, van Schaik W, Manson McGuire A, *et al.* Emergence of Epidemic Multidrug-Resistant Enterococcus faecium from Animal and Commensal Strains. *mBio* 2013; **4**: e00534-13.

210. Salmond GPC, Fineran PC. A century of the phage: past, present and future. *Nat Rev Microbiol* 2015; **13**: 777–86.

211. Davies EV, James CE, Williams D, *et al.* Temperate phages both mediate and drive adaptive evolution in pathogen biofilms. *Proc Natl Acad Sci* 2016; **113**: 8266–71.

212. Johnson CN, Sheriff EK, Duerkop BA, *et al.* Let me upgrade you: Impact of mobile genetic elements on enterococcal adaptation and evolution. *J Bacteriol* 2021; **203**: e0017721.

213. Duerkop B a, Palmer KL, Horsburgh MJ. Enterococcal Bacteriophages and Genome Defense. In: Gilmore MS, Clewell DB, Ike Y, et al., eds. *Enterococci: From Commensals to Leading Causes of Drug Resistant Infection [Internet]*. Boston: Massachusetts Eye and Ear Infirmary, 2014.

214. Mazaheri Nezhad Fard R, Barton MD, Heuzenroeder MW. Bacteriophage-mediated transduction of antibiotic resistance in enterococci. *Lett Appl Microbiol* 2011; **52**: 559–64.

215. Bennett PM. Genome plasticity: insertion sequence elements, transposons and integrons, and DNA rearrangement. *Methods Mol Biol* 2004; **266**: 71–113.

216. Werner G, Fleige C, Geringer U, *et al.* IS element IS16 as a molecular screening tool to identify hospital-associated strains of Enterococcus faecium. *BMC Infect Dis* 2011; **11**: 80.

217. Kirsch JM, Ely S, Stellfox ME, *et al.* Targeted IS-element sequencing uncovers transposition dynamics during selective pressure in enterococci. *PLOS Pathog* 2023; **19**: e1011424.

218. Salyers AA, Shoemaker NB, Stevens AM, *et al.* Conjugative transposons: an unusual and diverse set of integrated gene transfer elements. *Microbiol Rev* 1995; **59**: 579–90.

219. Bender JK, Fleige C, Klare I, *et al.* Detection of a cfr(B) Variant in German Enterococcus faecium Clinical Isolates and the Impact on Linezolid Resistance in Enterococcus spp. Rohde H, ed. *PLOS ONE* 2016; **11**: e0167042.

220. Zhu X-Q, Wang X-M, Li H, *et al.* Novel lnu (G) gene conferring resistance to lincomycin by nucleotidylation, located on Tn 6260 from Enterococcus faecalis E531. *J Antimicrob Chemother* 2017; **72**: 993–7.

221. Werner G, Coque TM, Hammerum AM, *et al.* Emergence and spread of vancomycin resistance among Enterococci in Europe. *Eurosurveillance* 2008; **13**: 1–11.

222. Clewell D, Weaver K, Dunny G, *et al.* Extrachromosomal and Mobile Elements in Enterococci: Transmission, Maintenance, and Epidemiology. In: Gilmore MS, Clewell DB, Ike Y, et al., eds. *Enterococci: From Commensals to Leading Causes of Drug Resistant Infection [Internet]*. Boston: Massachusetts Eye and Ear Infirmary, 2014. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24649510.

223. 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* 2002; **46**: 1410–6.

224. Jensen LB, Garcia-Migura L, Valenzuela AJS, *et al.* A classification system for plasmids from enterococci and other Gram-positive bacteria. *J Microbiol Methods* 2010; **80**: 25–43.

225. Arredondo-Alonso S, Top J, McNally A, *et al.* Plasmids Shaped the Recent Emergence of the Major Nosocomial Pathogen Enterococcus faecium. *mBio* 2020; **11**: e03284-19.

226. Hirt H, Manias DA, Bryan EM, *et al.* Characterization of the Pheromone Response of the *Enterococcus faecalis* Conjugative Plasmid pCF10: Complete Sequence and

Comparative Analysis of the Transcriptional and Phenotypic Responses of pCF10-Containing Cells to Pheromone Induction. *J Bacteriol* 2005; **187**: 1044–54.

227. Dunny GM. The peptide pheromone-inducible conjugation system of *Enterococcus faecalis* plasmid pCF10: cell–cell signalling, gene transfer, complexity and evolution. *Philos Trans R Soc B Biol Sci* 2007; **362**: 1185–93.

228. Lanza VF, Tedim AP, Martínez JL, *et al.* The Plasmidome of Firmicutes: Impact on the Emergence and the Spread of Resistance to Antimicrobials. *Microbiol Spectr* 2015; **3**: PLAS-0039-2014.

229. Hawkey J, Cottingham H, Tokolyi A, *et al.* Linear plasmids in Klebsiella and other Enterobacteriaceae. *Microb Genomics* 2022; **8**: 000807.

230. Hashimoto Y, Suzuki M, Kobayashi S, *et al.* Enterococcal Linear Plasmids Adapt to Enterococcus faecium and Spread within Multidrug-Resistant Clades. *Antimicrob Agents Chemother* 2023; **67**: e01619-22.

231. Gawryszewska I, Malinowska K, Kuch A, *et al.* Distribution of antimicrobial resistance determinants, virulence-associated factors and clustered regularly interspaced palindromic repeats loci in isolates of Enterococcus faecalis from various settings and genetic lineages. *Pathog Dis* 2017; **75**: 821–8.

232. Guzman Prieto AM, van Schaik W, Rogers MRC, *et al.* Global Emergence and Dissemination of Enterococci as Nosocomial Pathogens: Attack of the Clones? *Front Microbiol* 2016; **7**: 1–15.

233. Tedim AP, Ruiz-Garbajosa P, Corander J, *et al.* Population Biology of Intestinal Enterococcus Isolates from Hospitalized and Nonhospitalized Individuals in Different Age Groups. *Appl Environ Microbiol* 2015; **81**: 1811–22.

234. Willems RJL, Top J, Van Santen M, *et al.* Global spread of vancomycin-resistant Enterococcus faecium from distinct nosocomial genetic complex. *Emerg Infect Dis* 2005; **11**: 821–8.

235. Turner KME, Hanage WP, Fraser C, *et al.* Assessing the reliability of eBURST using simulated populations with known ancestry. *BMC Microbiol* 2007; **7**: 30.

236. Loman NJ, Pallen MJ. Twenty years of bacterial genome sequencing. *Nat Rev Microbiol* 2015; **13**: 787–94.

237. van Schaik W, Willems RJL. Genome-based insights into the evolution of enterococci. *Clin Microbiol Infect* 2010; **16**: 527–32.

238. Sahm DF, Kissinger J, Gilmore MS, *et al.* In Vitro Susceptibility Studies of Vancomycin-Resistant Enterococcus faecalis. *Antimicrob Agents Chemother* 1989; **33**: 1588–91.

239. Bourgogne A, Garsin DA, Qin X, *et al.* Large scale variation in Enterococcus faecalis illustrated by the genome analysis of strain OG1RF. *Genome Biol* 2008; **9**: R110.

240. Domann E, Hain T, Ghai R, *et al.* Comparative genomic analysis for the presence of potential enterococcal virulence factors in the probiotic Enterococcus faecalis strain Symbioflor 1. *Int J Med Microbiol* 2007; **297**: 533–9.

241. Qin X, Galloway-Peña JR, Sillanpaa J, *et al.* Complete genome sequence of Enterococcus faecium strain TX16 and comparative genomic analysis of Enterococcus faecium genomes. *BMC Microbiol* 2012; **12**: 135.

242. Palmer KL, Carniol K, Manson JM, *et al.* High-Quality Draft Genome Sequences of 28 Enterococcus sp. Isolates. *J Bacteriol* 2010; **192**: 2469–70.

243. van Schaik W, Top J, Riley DR, *et al.* Pyrosequencing-based comparative genome analysis of the nosocomial pathogen Enterococcus faecium and identification of a large transferable pathogenicity island. *BMC Genomics* 2010; **11**: 239.

244. Palmer KL, Godfrey P, Griggs A, *et al.* Comparative Genomics of Enterococci: Variation in Enterococcus faecalis, Clade Structure in E. faecium, and Defining Characteristics of E. gallinarum and E. casseliflavus. *mBio* 2012; **3**: e00318-11.

245. Galloway-Peña J, Roh JH, Latorre M, *et al.* Genomic and SNP Analyses Demonstrate a Distant Separation of the Hospital and Community-Associated Clades of Enterococcus faecium. *PLoS ONE* 2012; **7**: e30187.

246. de Been M, van Schaik W, Cheng L, *et al.* Recent Recombination Events in the Core Genome Are Associated with Adaptive Evolution in Enterococcus faecium. *Genome Biol Evol* 2013; **5**: 1524–35.

247. de Regt MJA, van Schaik W, van Luit-Asbroek M, *et al.* Hospital and community ampicillin-resistant Enterococcus faecium are evolutionarily closely linked but have diversified through niche adaptation. *PloS One* 2012; **7**: e30319.

248. Lebreton F, Manson AL, Saavedra JT, *et al.* Tracing the Enterococci from Paleozoic Origins to the Hospital. *Cell* 2017; **169**: 849–61.

249. Raven KE, Gouliouris T, Brodrick H, *et al.* Complex Routes of Nosocomial Vancomycin-Resistant Enterococcus faecium Transmission Revealed by Genome Sequencing. *Clin Infect Dis* 2017; **64**: 886–93.

250. Raven KE, Reuter S, Gouliouris T, *et al.* Genome-based characterization of hospital-adapted *Enterococcus faecalis* lineages. *Nat Microbiol* 2016; **1**: 15033.

251. Willems RJL, Top J, van Schaik W, *et al.* Restricted gene flow among hospital subpopulations of Enterococcus faecium. *mBio* 2012; **3**: e00151-12.

252. van Hal SJ, Willems RJL, Gouliouris T, *et al.* The interplay between community and hospital Enterococcus faecium clones within health-care settings: a genomic analysis. *Lancet Microbe* 2022; **3**: e133–41.

253. Belloso Daza MV, Cortimiglia C, Bassi D, *et al.* Genome-based studies indicate that the Enterococcus faecium Clade B strains belong to Enterococcus lactis species and lack of the hospital infection associated markers. *Int J Syst Evol Microbiol* 2021; **71**: 004948.

254. Pinholt M, Larner-Svensson H, Littauer P, *et al.* Multiple hospital outbreaks of vanA Enterococcus faecium in Denmark, 2012-13, investigated by WGS, MLST and PFGE. *J Antimicrob Chemother* 2015; **70**: 1–9.

255. van Hal SJ, Espedido BA, Coombs GW, *et al.* Polyclonal emergence of vanA vancomycin-resistant Enterococcus faecium in Australia. *J Antimicrob Chemother* 2017; **72**: 998–1001.

256. Landerslev KG, Jakobsen L, Olsen SS, *et al.* Polyclonal spread of vanA *Enterococcus faecium* in Central Denmark Region, 2009-2013, investigated using PFGE, MLST and WGS. *Int J Antimicrob Agents* 2016; **48**: 767–8.

257. Hammerum AM, Baig S, Kamel Y, *et al.* Emergence of vanA Enterococcus faecium in Denmark, 2005–15. *J Antimicrob Chemother* 2017; **72**: 2184–90.

258. Pinholt M, Gumpert H, Bayliss S, *et al.* Genomic analysis of 495 vancomycin-resistant Enterococcus faecium reveals broad dissemination of a vanA plasmid in more than 19 clones from Copenhagen, Denmark. *J Antimicrob Chemother* 2017; **72**: 40–7.

259. Bender JK, Kalmbach A, Fleige C, *et al.* Population structure and acquisition of the vanB resistance determinant in German clinical isolates of Enterococcus faecium ST192. *Sci Rep* 2016; **6**: 21847.

260. Howden BP, Holt KE, Lam MMC, *et al.* Genomic Insights to Control the Emergence of Vancomycin-Resistant Enterococci. *mBio* 2013; **4**: e00412-13.

261. Sivertsen A, Billström H, Melefors Ö, *et al.* A Multicentre Hospital Outbreak in Sweden Caused by Introduction of a vanB2 Transposon into a Stably Maintained pRUM-Plasmid in an Enterococcus faecium ST192 Clone van Schaik W, ed. *PLoS ONE* 2014; **9**: e103274.

262. Reuter S, Ellington MJ, Cartwright EJP, *et al.* Rapid Bacterial Whole-Genome Sequencing to Enhance Diagnostic and Public Health Microbiology. *JAMA Intern Med* 2013; **173**: 1397–404.

263. Salipante SJ, SenGupta DJ, Cummings LA, *et al.* Application of Whole-Genome Sequencing for Bacterial Strain Typing in Molecular Epidemiology. *J Clin Microbiol* 2015; **53**: 1072–9.

264. Lytsy B, Engstrand L, Gustafsson Å, *et al.* Time to review the gold standard for genotyping vancomycin-resistant enterococci in epidemiology: Comparing whole-genome sequencing with PFGE and MLST in three suspected outbreaks in Sweden during 2013–2015. *Infect Genet Evol* 2017; **54**: 74–80.

265. Pöntinen AK, Top J, Arredondo-Alonso S, *et al.* Apparent nosocomial adaptation of Enterococcus faecalis predates the modern hospital era. *Nat Commun* 2021; **12**: 1523.

266. Lister DM, Kotsanas D, Ballard SA, *et al.* Outbreak of vanB vancomycin-resistant Enterococcus faecium colonization in a neonatal service. *Am J Infect Control* 2015; **43**: 1061–5.

267. Brodrick HJ, Raven KE, Harrison EM, *et al.* Whole-genome sequencing reveals transmission of vancomycin-resistant Enterococcus faecium in a healthcare network. *Genome Med* 2016; **8**: 4.

268. Bashir A, Attie O, Sullivan M, *et al.* Genomic confirmation of vancomycin-resistant Enterococcus transmission from deceased donor to liver transplant recipient. *PloS One* 2017; **12**: e0170449.

269. EUCAST. Breakpoint tables for interpretation of MICs and zone diameters, Version 13.0. 2023. Available at: http://www.eucast.org.

270. EUCAST. Breakpoint tables for interpretation of MICs and zone diameters, Version 8.0. 2018. Available at: http://www.eucast.org. Accessed March 1, 2018.

271. Woodford N, Tysall L, Auckland C, *et al.* Detection of oxazolidinone-resistant *Enterococcus faecalis* and *Enterococcus faecium* strains by real-time PCR and PCR-restriction fragment length polymorphism analysis. *J Clin Microbiol* 2002; **40**: 4298–300.

272. Werner G, Strommenger B, Klare I, *et al.* Molecular detection of linezolid resistance in *Enterococcus faecium* and *Enterococcus faecalis* by use of 5' nuclease real-time PCR compared to a modified classical approach. *J Clin Microbiol* 2004; **42**: 5327–31.

273. Kehrenberg C, Schwarz S. Distribution of florfenicol resistance genes *fexA* and *cfr* among chloramphenicol-resistant *Staphylococcus* isolates. *Antimicrob Agents Chemother* 2006; **50**: 1156–63.

274. Morrison D, Woodford N, Barrett SP, *et al.* DNA Banding Pattern Polymorphism in Vancomycin-Resistant Enterococcus faecium and Criteria for Defining Strains. *J Clin Microbiol* 1999; **37**: 1084–91.

275. Loman NJ, Quinlan AR. Poretools: A toolkit for analyzing nanopore sequence data. *Bioinformatics* 2014; **30**: 3399–401.

276. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinforma Oxf Engl* 2014; **30**: 2114–20.

277. Ewels P, Magnusson M, Lundin S, *et al.* MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 2016; **32**: 3047–8.

278. Jolley KA, Maiden MC. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* 2010; **11**: 595.

279. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications [version 1; peer review: 2 approved]. *Wellcome Open Res* 2018; **3**: 124.

280. Carver T, Harris SR, Berriman M, *et al.* Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinforma Oxf Engl* 2012; **28**: 464–9.

281. Page AJ, Taylor B, Delaney AJ, *et al.* SNP-sites: Rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb Genomics* 2016; **2**: e000056.

282. Moradigaravand D, Gouliouris T, Blane B, *et al.* Within-host evolution of Enterococcus faecium during longitudinal carriage and transition to bloodstream infection in immunocompromised patients. *Genome Med* 2017; **9**: 119.

283. van Hal SJ, Willems RJL, Gouliouris T, *et al.* The global dissemination of hospital clones of Enterococcus faecium. *Genome Med* 2021; **13**: 52.

284. Lemonidis K, Salih TS, Dancer SJ, *et al.* Emergence of an Australian-like pstS-null vancomycin resistant Enterococcus faecium clone in Scotland. *PLOS ONE* 2019; **14**: e0218185.

285. Pinholt M, Bayliss SC, Gumpert H, *et al.* WGS of 1058 Enterococcus faecium from Copenhagen, Denmark, reveals rapid clonal expansion of vancomycin-resistant clone ST80 combined with widespread dissemination of a vanA-containing plasmid and acquisition of a heterogeneous accessory genome. *J Antimicrob Chemother* 2019; **74**: 1776–85.

286. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* 2013; **30**: 772–80.

287. Hoang DT, Chernomor O, von Haeseler A, *et al.* UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol* 2018; **35**: 518–22.

288. Kalyaanamoorthy S, Minh BQ, Wong TKF, *et al.* ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017; **14**: 587–9.

289. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R Schwartz R, ed. *Bioinformatics* 2019; **35**: 526–8.

290. R Core Team. R: A Language and Environment for Statistical Computing. 2023. Available at: https://www.R-project.org/.

291. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 2021; **49**: W293–6.

292. Pruitt KD, Tatusova T, Brown GR, *et al.* NCBI Reference Sequences (RefSeq): Current status, new features and genome annotation policy. *Nucleic Acids Res* 2012; **40**: D130–5.

293. De Coster W, D'Hert S, Schultz DT, *et al.* NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 2018; **34**: 2666–9.

294. Ruan J, Li H. Fast and accurate long-read assembly with wtdbg2. *Nat Methods* 2020; **17**: 155–8.

295. Vaser R, Šikić M. Time- and memory-efficient genome assembly with Raven. *Nat Comput Sci* 2021; **1**: 332–6.

296. Simão FA, Waterhouse RM, Ioannidis P, *et al.* BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 2015; **31**: 3210–2.

297. Zankari E, Allesøe R, Joensen KG, *et al.* PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. *J Antimicrob Chemother* 2017; **72**: 2764–8.

298. Arndt D, Grant JR, Marcu A, *et al.* PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 2016; **44**: W16-21.

299. Ondov BD, Treangen TJ, Melsted P, *et al.* Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol* 2016; **17**: 132.

300. Bourgeois-Nicolaos N, Massias L, Couson B, *et al.* Dose dependence of emergence of resistance to linezolid in Enterococcus faecalis in vivo. *J Infect Dis* 2007; **195**: 1480–8.

301. Ghosh A, N S, Saha S. Survey of drug resistance associated gene mutations in Mycobacterium tuberculosis, ESKAPE and other bacterial species. *Sci Rep* 2020; **10**: 8957.

302. Zankari E, Hasman H, Kaas RS, *et al.* Genotyping using whole-genome sequencing is a realistic alternative to surveillance based on phenotypic antimicrobial susceptibility testing. *J Antimicrob Chemother* 2013; **68**: 771–7.

303. FDA. *Antimicrobial Susceptibility Test (AST) Systems - Class II Special Controls Guidance for Industry and FDA*. Food and Drug Administration; 2009. Available at: https://www.fda.gov/medical-devices/guidance-documents-medical-devices-and-radiation-emitting-products/antimicrobial-susceptibility-test-ast-systems-class-ii-special-controls-guidance-industry-and-fda. Accessed June 29, 2023.

304. Wymant C, Hall M, Ratmann O, *et al.* PHYLOSCANNER: Inferring Transmission from Within- and Between-Host Pathogen Genetic Diversity. *Mol Biol Evol* 2018; **35**: 719–33.

305. Higgs C, Sherry NL, Seemann T, *et al.* Optimising genomic approaches for identifying vancomycin-resistant Enterococcus faecium transmission in healthcare settings. *Nat Commun* 2022; **13**: 509.

306. Shannon P, Markiel A, Ozier O, *et al.* Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res* 2003; **13**: 2498–504.

307. Prjibelski A, Antipov D, Meleshko D, *et al.* Using SPAdes De Novo Assembler. *Curr Protoc Bioinforma* 2020; **70**: e102.

308. Didelot X, Croucher NJ, Bentley SD, *et al.* Bayesian inference of ancestral dates on bacterial phylogenetic trees. *Nucleic Acids Res* 2018; **46**: e134.

309. Didelot X, Siveroni I, Volz EM. Additive Uncorrelated Relaxed Clock Models for the Dating of Genomic Epidemiology Phylogenies. *Mol Biol Evol* 2021; **38**: 307–17.

310. Pidot SJ, Gao W, Buultjens AH, *et al.* Increasing tolerance of hospital Enterococcus faecium to handwash alcohols. *Sci Transl Med* 2018; **10**: eaar6115.

311. Reuter JS, Mathews DH. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* 2010; **11**: 129.

312. Sullivan MJ, Petty NK, Beatson SA. Easyfig: A genome comparison visualizer. *Bioinformatics* 2011; **27**: 1009–10.

313. Alikhan N-F, Petty NK, Ben Zakour NL, *et al.* BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 2011; **12**: 402.

314. Huebner RE, Dagan R, Porath N, *et al.* Lack of Utility of Serotyping Multiple Colonies for Detection of Simultaneous Nasopharyngeal Carriage of Different Pneumococcal Serotypes. *Pediatr Infect Dis J* 2000; **19**: 1017–20.

315. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York; 2016. Available at: https://ggplot2.tidyverse.org.

316. Pedersen TL. *patchwork: The Composer of Plots*. 2020. Available at: https://CRAN.R-project.org/package=patchwork.

317. Wei T, Simko V. *R package 'corrplot': Visualization of a Correlation Matrix*. 2021. Available at: https://github.com/taiyun/corrplot.

318. Warnes G, Bolker B, Bonebakker L, *et al. gplots: Various R Programming Tools for Plotting Data*. 2022. Available at: https://CRAN.R-project.org/package=gplots.

319. Kassambara A. *rstatix: Pipe-Friendly Framework for Basic Statistical Tests*. 2023. Available at: https://CRAN.R-project.org/package=rstatix.

320. Public Health England. *English surveillance programme for antimicrobial utilisation and resistance (ESPAUR) Report 2018-2019*. London, UK: PHE; 2019. Available at: https://www.gov.uk/government/publications/english-surveillance-programme-antimicrobial-utilisation-and-resistance-espaur-report.

321. Mendes RE, Deshpande LM, Jones RN. Linezolid update: Stable *in vitro* activity following more than a decade of clinical use and summary of associated resistance mechanisms. *Drug Resist Updat* 2014; **17**: 1–12.

322. Abbo L, Shukla BS, Giles A, *et al.* Linezolid and vancomycin-resistant *Enterococcus faecium* in solid organ transplant recipients: Infection control and antimicrobial stewardship using whole genome sequencing. *Clin Infect Dis* 2019; **69**: 259–65.

323. Diaz L, Kiratisin P, Mendes RE, *et al.* Transferable Plasmid-Mediated Resistance to Linezolid Due to cfr in a Human Clinical Isolate of Enterococcus faecalis. *Antimicrob Agents Chemother* 2012; **56**: 3917–22.

324. Pang S, Boan P, Lee T, *et al.* Linezolid-resistant ST872 *Enteroccocus faecium* harbouring *optrA* and *cfr* (D) oxazolidinone resistance genes. *Int J Antimicrob Agents* 2020; **55**: 105831.

325. Long KS, Poehlsgaard J, Kehrenberg C, *et al.* The Cfr rRNA methyltransferase confers resistance to phenicols, lincosamides, oxazolidinones, pleuromutilins, and streptogramin A antibiotics. *Antimicrob Agents Chemother* 2006; **50**: 2500–5.

326. Cai J, Wang Y, Schwarz S, *et al.* High detection rate of the oxazolidinone resistance gene *optrA* in *Enterococcus faecalis* isolated from a Chinese anorectal surgery ward. *Int J Antimicrob Agents* 2016; **48**: 757–9.

327. He T, Shen Y, Schwarz S, *et al.* Genetic environment of the transferable oxazolidinone/phenicol resistance gene *optrA* in *Enterococcus faecalis* isolates of human and animal origin. *J Antimicrob Chemother* 2016; **71**: 1466–73.

328. Goodhead I, Darby AC. Taking the pseudo out of pseudogenes. *Curr Opin Microbiol* 2015; **23**: 102–9.

329. Zou J, Tang Z, Yan J, *et al.* Dissemination of Linezolid Resistance Through Sex Pheromone Plasmid Transfer in Enterococcus faecalis. *Front Microbiol* 2020; **11**.

330. Freitas AR, Finisterra L, Tedim AP, *et al.* Linezolid- and Multidrug-Resistant Enterococci in Raw Commercial Dog Food, Europe, 2019–2020 - Volume 27, Number 8—August 2021 - Emerging Infectious Diseases journal - CDC. 2021. Available at: https://wwwnc.cdc.gov/eid/article/27/8/20-4933_article. Accessed November 23, 2021.

331. Limura M, Hayashi W, Arai E, *et al.* Identification of a multiresistant mosaic plasmid carrying a new segment of IS1216E-flanked optrA with integrated Tn551-ermB element in linezolid-resistant Enterococcus faecalis human isolate. *J Glob Antimicrob Resist* 2020; **22**: 697–9.

332. Egan SA, Shore AC, O'Connell B, *et al.* Linezolid resistance in Enterococcus faecium and Enterococcus faecalis from hospitalized patients in Ireland: high prevalence of the MDR genes optrA and poxtA in isolates with diverse genetic backgrounds. *J Antimicrob Chemother* 2020; **75**: 1704–11.

333. Partridge SR, Kwong SM, Firth N, *et al.* Mobile genetic elements associated with antimicrobial resistance. *Clin Microbiol Rev* 2018; **31**: e00088-17.

334. Tansirichaiya S, Rahman MdA, Roberts AP. The Transposon Registry. *Mob DNA* 2019; **10**: 40.

335. Li D, Li X-Y, Schwarz S, *et al.* Tn *6674* , a Novel Enterococcal *optrA* -Carrying Multiresistance Transposon of the Tn *554* Family. *Antimicrob Agents Chemother* 2019; **63**: e00809-29.

336. Freitas AR, Tedim AP, Novais C, *et al.* Comparative genomics of global optrA-carrying Enterococcus faecalis uncovers a common chromosomal hotspot for optrA acquisition within a diversity of core and accessory genomes. *Microb Genomics* 2020; **6**: e000350.

337. Chen L, Han D, Tang Z, *et al.* Co-existence of the oxazolidinone resistance genes cfr and optrA on two transferable multi-resistance plasmids in one Enterococcus faecalis isolate from swine. *Int J Antimicrob Agents* 2020; **56**: 105993.

338. Larsen J, Schønheyder HC, Lester CH, *et al.* Porcine-origin gentamicin-resistant *Enterococcus faecalis* in humans, Denmark. *Emerg Infect Dis* 2010; **16**: 682–4.

339. Shang Y, Li D, Hao W, *et al.* A prophage and two ICESa2603-family integrative and conjugative elements (ICEs) carrying optrA in Streptococcus suis. *J Antimicrob Chemother* 2019; **74**: 2876–9.

340. Hao W, Shan X, Li D, *et al.* Analysis of a poxtA- and optrA-co-carrying conjugative multiresistance plasmid from Enterococcus faecalis. *J Antimicrob Chemother* 2019; **74**: 1771–5.

341. Cavaco LM, Bernal JF, Zankari E, *et al.* Detection of linezolid resistance due to the *optrA* gene in *Enterococcus faecalis* from poultry meat from the American continent (Colombia). *J Antimicrob Chemother* 2017; **72**: 678–83.

342. Cavaco LM, Korsgaard H, Kaas RS, *et al.* First detection of linezolid resistance due to the *optrA* gene in enterococci isolated from food products in Denmark. *J Glob Antimicrob Resist* 2017; **9**: 128–9.

343. Elghaieb H, Freitas AR, Abbassi MS, *et al.* Dispersal of linezolid-resistant enterococci carrying *poxtA* or *optrA* in retail meat and food-producing animals from Tunisia. *J Antimicrob Chemother* 2019; **75**: 2865–9.

344. Wu Y, Fan R, Wang Y, *et al.* Analysis of combined resistance to oxazolidinones and phenicols among bacteria from dogs fed with raw meat/vegetables and the respective food items. *Sci Rep* 2019; **9**: 15500.

345. Elghaieb H, Tedim AP, Abbassi MS, *et al.* From farm to fork: identical clones and Tn6674-like elements in linezolid-resistant Enterococcus faecalis from food-producing animals and retail meat. *J Antimicrob Chemother* 2020; **75**: 30–5.

346. Shen W, Zhang R, Cai J. Co-occurrence of multiple plasmid-borne linezolid resistance genes—optrA, cfr, poxtA2 and cfr(D) in an Enterococcus faecalis isolate from retail meat. *J Antimicrob Chemother* 2023; **78**: 1637–43.

347. Zarzecka U, Zakrzewski AJ, Chajęcka-Wierzchowska W, *et al.* Linezolid-Resistant Enterococcus spp. Isolates from Foods of Animal Origin-The Genetic Basis of Acquired Resistance. *Foods* 2022; **11**: 975.

348. Freitas AR, Finisterra L, Tedim AP, *et al.* Linezolid- and Multidrug-Resistant Enterococci in Raw Commercial Dog Food, Europe, 2019-2020. *Emerg Infect Dis* 2021; **27**: 2221–4.

349. Nüesch-Inderbinen M, Heyvaert L, Treier A, *et al.* High occurrence of Enterococcus faecalis, Enterococcus faecium, and Vagococcus lutrae harbouring oxazolidinone resistance genes in raw meat-based diets for companion animals - a public health issue, Switzerland, September 2018 to May 2020. *Eurosurveillance* 2023; **28**: 2200496.

350. Nüesch-Inderbinen M, Biggel M, Zurfluh K, *et al.* Faecal carriage of enterococci harbouring oxazolidinone resistance genes among healthy humans in the community in Switzerland. *J Antimicrob Chemother* 2022; **77**: 2779–83.

351. Cai J, Schwarz S, Chi D, *et al.* Faecal carriage of *optrA*-positive enterococci in asymptomatic healthy humans in Hangzhou, China. *Clin Microbiol Infect* 2019; **25**: 630.e1-630.e6.

352. Koutsoumanis K, Allende A, Álvarez-Ordóñez A, *et al.* Role played by the environment in the emergence and spread of antimicrobial resistance (AMR) through the food chain. *EFSA J* 2021; **19**: e06651.

353. Munk P, Knudsen BE, Lukjancenko O, *et al.* Abundance and diversity of the faecal resistome in slaughter pigs and broilers in nine European countries. *Nat Microbiol* 2018; **3**: 898–908.

354. Zhao Q, Wang Y, Wang S, *et al.* Prevalence and abundance of florfenicol and linezolid resistance genes in soils adjacent to swine feedlots. *Sci Rep* 2016; **6**: 1–7.

355. Sassi M, Guerin F, Zouari A, *et al.* Emergence of *optrA*-mediated linezolid resistance in enterococci from France, 2006-2016. *J Antimicrob Chemother* 2019; **74**: 1469–72.

356. Bender JK, Fleige C, Lange D, *et al.* Rapid emergence of highly variable and transferable oxazolidinone and phenicol resistance gene *optrA* in German *Enterococcus* spp. clinical isolates. *Int J Antimicrob Agents* 2018; **52**: 819–27.

357. Egan SA, Corcoran S, McDermott H, *et al.* Hospital outbreak of linezolid-resistant and vancomycin-resistant ST80 Enterococcus faecium harbouring an optrA-encoding conjugative plasmid investigated by whole-genome sequencing. *J Hosp Infect* 2020; **105**: 726–35.

358. Lazaris A, Coleman DC, Kearns AM, *et al.* Novel multiresistance cfr plasmids in linezolid-resistant methicillin-resistant Staphylococcus epidermidis and vancomycin-

resistant Enterococcus faecium (VRE) from a hospital outbreak: co-location of cfr and optrA in VRE. *J Antimicrob Chemother* 2017; **72**: 3252–7.

359. Zhou X, Willems RJL, Friedrich AW, *et al.* Enterococcus faecium: from microbiological insights to practical recommendations for infection control and diagnostics. *Antimicrob Resist Infect Control* 2020; **9**: 130.

360. Croucher NJ, Didelot X. The application of genomics to tracing bacterial pathogen transmission. *Curr Opin Microbiol* 2015; **23**: 62–7.

361. Bryant JM, Grogono DM, Greaves D, *et al.* Whole-genome sequencing to identify transmission of Mycobacterium abscessus between patients with cystic fibrosis: a retrospective cohort study. *The Lancet* 2013; **381**: 1551–60.

362. Worby CJ, Lipsitch M, Hanage WP. Within-Host Bacterial Diversity Hinders Accurate Reconstruction of Transmission Networks from Genomic Distance Data. *PLoS Comput Biol* 2014; **10**: e1003549.

363. Snitkin ES, Zelazny AM, Thomas PJ, *et al.* Tracking a Hospital Outbreak of Carbapenem-Resistant Klebsiella pneumoniae with Whole-Genome Sequencing. *Sci Transl Med* 2012; **4**: 148ra116.

364. Hall MD, Holden MT, Srisomang P, *et al.* Improved characterisation of MRSA transmission using within-host bacterial sequence diversity. *eLife* 2019; **8**: e46402.

365. Harris SR, Cartwright EJ, Török ME, *et al.* Whole-genome sequencing for analysis of an outbreak of meticillin-resistant Staphylococcus aureus: a descriptive study. *Lancet Infect Dis* 2013; **13**: 130–6.

366. Okoro CK, Kingsley RA, Quail MA, *et al.* High-Resolution Single Nucleotide Polymorphism Analysis Distinguishes Recrudescence and Reinfection in Recurrent Invasive Nontyphoidal Salmonella Typhimurium Disease. *Clin Infect Dis* 2012; **54**: 955–63.

367. Paterson GK, Harrison EM, Murray GGR, *et al.* Capturing the cloud of diversity reveals complexity and heterogeneity of MRSA carriage, infection and transmission. *Nat Commun* 2015; **6**: 6560.

368. Dubin KA, Mathur D, McKenney PT, *et al.* Diversification and evolution of vancomycin-resistant Enterococcus faecium during intestinal domination. *Infect Immun* 2019; **87**: e00102-19.

369. Bayjanov JR, Baan J, Rogers MRC, *et al.* Enterococcus faecium genome dynamics during long-term asymptomatic patient gut colonization. *Microb Genomics* 2019; **5**: e000277.

370. Both A, Kruse F, Mirwald N, *et al.* Population dynamics in colonizing vancomycin-resistant E. faecium isolated from immunosuppressed patients. *J Glob Antimicrob Resist* 2022; **28**: 267–73.

371. Lebreton F, Schaik W van, McGuire AM, *et al.* Emergence of Epidemic Multidrug-Resistant Enterococcus faecium from Animal and Commensal Strains. *mBio* 2013; **4**: e00534-13.

372. Dingle KE, Elliott B, Robinson E, *et al.* Evolutionary History of the Clostridium difficile Pathogenicity Locus. *Genome Biol Evol* 2014; **6**: 36–52.

373. Hashimoto Y, Taniguchi M, Uesaka K, *et al.* Novel Multidrug-Resistant Enterococcal Mobile Linear Plasmid pELF1 Encoding vanA and vanM Gene Clusters From a Japanese Vancomycin-Resistant Enterococci Isolate. *Front Microbiol* 2019; **10**: 2568.

374. Boumasmoud M, Dengler Haunreiter V, Schweizer TA, *et al.* Genomic Surveillance of Vancomycin-Resistant Enterococcus faecium Reveals Spread of a Linear Plasmid Conferring a Nutrient Utilization Advantage. *mBio* 2022; **13**: e03771-21.

375. Gouliouris T, Coll F, Ludden C, *et al.* Quantifying acquisition and transmission of Enterococcus faecium using genomic surveillance. *Nat Microbiol* 2021; **6**: 103–11.

376. Gouliouris T, Coll F, Ludden C, *et al.* Quantifying acquisition and transmission of Enterococcus faecium using genomic surveillance. *Nat Microbiol* 2021; **6**: 103–11.

377. Mäklin T, Kallonen T, David S, *et al.* High-resolution sweep metagenomics using fast probabilistic inference [version 2; peer review: 2 approved]. *Wellcome Open Res* 2021; **5**: 14.

378. Anyansi C, Straub TJ, Manson AL, *et al.* Computational Methods for Strain-Level Microbial Detection in Colony and Metagenome Sequencing Data. *Front Microbiol* 2020; **11**: 1925.

379. Mu A, Kwong JC, Isles NS, *et al.* Reconstruction of the Genomes of Drug-Resistant Pathogens for Outbreak Investigation through Metagenomic Sequencing. *mSphere* 2019; **4**: e00529-18.

380. Xanthopoulou K, Wille J, Zweigner J, *et al.* Characterization of a vancomycin-resistant Enterococcus faecium isolate and a vancomycin-susceptible E. faecium isolate from the same blood culture. *J Antimicrob Chemother* 2020; **76**: 883–6.

381. Cárdenas AM, Andreacchio KA, Edelstein PH. Prevalence and detection of mixed-population enterococcal bacteremia. *J Clin Microbiol* 2014; **52**: 2604–8.

382. McInnes RS, Snaith AE, Dunn SJ, *et al.* A novel resistance reversion mechanism in a vancomycin-variable Enterococcus faecium strain. 2023. Available at: https://www.biorxiv.org/content/10.1101/2023.04.28.538149v1. Accessed June 14, 2023.

383. Smith AB, Jenior ML, Keenan O, *et al.* Enterococci enhance Clostridioides difficile pathogenesis. *Nature* 2022; **611**: 780–6.

384. Fujiya Y, Harada T, Sugawara Y, *et al.* Transmission dynamics of a linear vanA-plasmid during a nosocomial multiclonal outbreak of vancomycin-resistant enterococci in a non-endemic area, Japan. *Sci Rep* 2021; **11**: 14780.

385. Hashimoto Y, Kita I, Suzuki M, *et al.* First Report of the Local Spread of Vancomycin-Resistant Enterococci Ascribed to the Interspecies Transmission of a vanA Gene Cluster-Carrying Linear Plasmid. *mSphere* 2020; **5**: e00102-20.

386. Lee AS, White E, Monahan LG, *et al.* Defining the Role of the Environment in the Emergence and Persistence of vanA Vancomycin-Resistant Enterococcus (VRE) in an Intensive Care Unit: A Molecular Epidemiological Study. *Infect Control Hosp Epidemiol* 2018; **39**: 668–75.

387. Ford CD, Lopansri BK, Gazdik MA, *et al.* Room contamination, patient colonization pressure, and the risk of vancomycin-resistant Enterococcus colonization on a unit dedicated to the treatment of hematologic malignancies and hematopoietic stem cell transplantation. *Am J Infect Control* 2016; **44**: 1110–5.

388. McDermott H, Skally M, O'Rourke J, *et al.* Near-patient environmental contamination of an intensive care unit with Vancomycin-resistant Enterococci (VRE) and Extended-Spectrum Beta-Lactamase–Producing Enterobacteriaceae (ESBL-E) before and after the introduction of chlorhexidine bathing for patients. *Infect Control Hosp Epidemiol* 2018; **39**: 1131–2.

389. Gouliouris T, Blane B, Brodrick HJ, *et al.* Comparison of two chromogenic media for the detection of vancomycin-resistant enterococcal carriage by nursing home residents. *Diagn Microbiol Infect Dis* 2016; **85**: 409–12.

390. D'Agata EMC, Gautam S, Green WK, *et al.* High rate of false-negative results of the rectal swab culture method in detection of gastrointestinal colonization with vancomycin-resistant enterococci. *Clin Infect Dis* 2002; **34**: 167–72.

391. Linfield RY, Campeau S, Injean P, *et al.* Practical methods for effective vancomycin-resistant enterococci (VRE) surveillance: experience in a liver transplant surgical intensive care unit. *Infect Control Hosp Epidemiol* 2018; **39**: 1178–82.

392. Gordon LG, Elliott TM, Forde B, *et al.* Budget impact analysis of routinely using whole-genomic sequencing of six multidrug-resistant bacterial pathogens in Queensland, Australia. *BMJ Open* 2021; **11**: e041968.

393. Torok E, Brodrick H, Khokhar F, *et al.* Prospective Surveillance and Rapid Whole-Genome Sequencing Detects Two Unsuspected Outbreaks of Carbapenemase-Producing Klebsiella pneumoniae in a UK Teaching Hospital. *Open Forum Infect Dis* 2017; **4**: S43–4.

394. Sherry NL, Gorrie CL, Kwong JC, *et al.* Multi-site implementation of whole genome sequencing for hospital infection control: A prospective genomic epidemiological analysis. *Lancet Reg Health - West Pac* 2022; **23**: 100446.

395. Forde BM, Bergh H, Cuddihy T, *et al.* Clinical Implementation of Routine Whole-genome Sequencing for Hospital Infection Control of Multi-drug Resistant Pathogens. *Clin Infect Dis* 2023; **76**: e1277–84.

396. Fox JM, Saunders NJ, Jerwood SH. Economic and health impact modelling of a whole genome sequencing-led intervention strategy for bacterial healthcare-associated infections for England and for the USA. *Microb Genomics* 2023; **9**: 001087.

397. Weterings V, van Oosten A, Nieuwkoop E, *et al.* Management of a hospital-wide vancomycin-resistant Enterococcus faecium outbreak in a Dutch general hospital, 2014–2017: successful control using a restrictive screening strategy. *Antimicrob Resist Infect Control* 2021; **10**: 38.

398. Vuichard-Gysin D, Sommerstein R, Kronenberg A, *et al.* High adherence to national IPC guidelines as key to sustainable VRE control in Swiss hospitals: a cross-sectional survey. *Antimicrob Resist Infect Control* 2022; **11**: 19.

399. Escaut L, Bouam S, Frank-Soltysiak M, *et al.* Eradication of an outbreak of vancomycin-resistant Enterococcus (VRE): the cost of a failure in the systematic screening. *Antimicrob Resist Infect Control* 2013; **2**: 18.

400. van der Pol S, Lokate M, Postma MJ, *et al.* Costs of two vancomycin-resistant enterococci outbreaks in an academic hospital. *Antimicrob Steward Healthc Epidemiol* 2023; **3**: e8.

401. Maechler F, Weber A, Schwengers O, *et al.* Split k-mer analysis compared to cgMLST and SNP-based core genome analysis for detecting transmission of vancomycin-resistant enterococci: results from routine outbreak analyses across different hospitals and hospitals networks in Berlin, Germany. *Microb Genomics* 2023; **9**: 000937.

402. Tonkin-Hill G, Lees JA, Bentley SD, *et al.* Fast hierarchical Bayesian analysis of population structure. *Nucleic Acids Res* 2019; **47**: 5539–49.

403. Piezzi V, Wassilew N, Atkinson A, *et al.* Nosocomial outbreak of vancomycin-resistant Enterococcus faecium (VRE) ST796, Switzerland, 2017 to 2020. *Eurosurveillance* 2022; **27**: 2200285.

404. Zhou X, Chlebowicz MA, Bathoorn E, *et al.* Elucidating vancomycin-resistant Enterococcus faecium outbreaks: the role of clonal spread and movement of mobile genetic elements. *J Antimicrob Chemother* 2018; **73**: 3259–67.

405. Gunnink LB, Arouri DJ, Jolink FEJ, *et al.* Compliance to Screening Protocols for Multidrug-Resistant Microorganisms at the Emergency Departments of Two Academic Hospitals in the Dutch-German Cross-Border Region. *Trop Med Infect Dis* 2021; **6**: 15.

406. Martischang R, Buetti N, Balmelli C, *et al.* Nation-wide survey of screening practices to detect carriers of multi-drug resistant organisms upon admission to Swiss healthcare institutions. *Antimicrob Resist Infect Control* 2019; **8**: 37.

407. Permana B, Harris PNA, Runnegar N, *et al.* Using Genomics To Investigate an Outbreak of Vancomycin-Resistant Enterococcus faecium ST78 at a Large Tertiary Hospital in Queensland. *Microbiol Spectr* 2023: e0420422.

408. Cassone M, Wang J, Lansing BJ, *et al.* Diversity and Persistence of MRSA and VRE in Skilled Nursing Facilities: Environmental Screening, Whole Genome Sequencing, Development of a Dispersion Index. *J Hosp Infect* 2023; **138**: 8–18.

409. UKHSA. *English surveillance programme for antimicrobial utilisation and resistance (ESPAUR) Report 2021 to 2022*. London, UK: UK Health Security Agency; 2022. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1118310/ESPAUR-report-2021-to-2022.pdf.

410. Bender JK, Hermes J, Zabel LT, *et al.* Controlling an Unprecedented Outbreak with Vancomycin-Resistant Enterococcus faecium in Germany, October 2015 to November 2019. *Microorganisms* 2022; **10**: 1603.

411. Lee RS, Gonçalves da Silva A, Baines SL, *et al.* The changing landscape of vancomycin-resistant Enterococcus faecium in Australia: a population-level genomic study. *J Antimicrob Chemother* 2018; **73**: 3268–78.

412. Egan SA, Kavanagh NL, Shore AC, *et al.* Genomic analysis of 600 vancomycin-resistant Enterococcus faecium reveals a high prevalence of ST80 and spread of similar vanA regions via IS1216E and plasmid transfer in diverse genetic lineages in Ireland. *J Antimicrob Chemother* 2022; **77**: 320–30.

413. Brown AR, Amyes SG, Paton R, *et al.* Epidemiology and control of vancomycin-resistant enterococci (VRE) in a renal unit. *J Hosp Infect* 1998; **40**: 115–24.

414. Nelson RRS, McGregor KF, Brown AR, *et al.* Isolation and Characterization of Glycopeptide-Resistant Enterococci from Hospitalized Patients over a 30-Month Period. *J Clin Microbiol* 2000; **38**: 2112–6.

415. Inkster T, Coia J, Meunier D, *et al.* First outbreak of colonization by linezolid- and glycopeptide-resistant Enterococcus faecium harbouring the cfr gene in a UK nephrology unit. *J Hosp Infect* 2017; **97**: P397-402.

416. Health Protection Scotland. *Scottish One Health Antimicrobial Use and Antimicrobial Resistance Report 2016*. Glasgow, UK: Health Protection Scotland; 2017. Available at: https://www.nss.nhs.scot/media/3394/sonaar-2021-report.pdf.

417. Health Protection Scotland. *Scottish One Health Antimicrobial Use and Antimicrobial Resistance Report in 2017*. Glasgow, UK: Health Protection Scotland; 2018.

418. Misiakou M-A, Hertz FB, Schønning K, *et al.* Emergence of linezolid-resistant Enterococcus faecium in a tertiary hospital in Copenhagen. *Microb Genomics* 2023; **9**: 001055.

419. Leavis HL, Bonten MJ, Willems RJ. Identification of high-risk enterococcal clonal complexes: global dispersion and antibiotic resistance. *Curr Opin Microbiol* 2006; **9**: 454–60.

420. Tedim AP, Ruíz-Garbajosa P, Rodríguez MC, *et al.* Long-term clonal dynamics of Enterococcus faecium strains causing bloodstream infections (1995–2015) in Spain. *J Antimicrob Chemother* 2017; **72**: 48–55.

421. Werner G, Neumann B, Weber RE, *et al.* Thirty years of VRE in Germany – "expect the unexpected": The view from the National Reference Centre for Staphylococci and Enterococci. *Drug Resist Updat* 2020; **53**: 100732.

422. Nathwani D, Christie P. The Scottish approach to enhancing antimicrobial stewardship. *J Antimicrob Chemother* 2007; **60**: i69–71.

423. Nathwani D, Sneddon J, Malcolm W, *et al.* Scottish Antimicrobial Prescribing Group (SAPG): development and impact of the Scottish National Antimicrobial Stewardship Programme. *Int J Antimicrob Agents* 2011; **38**: 16–26.

424. MacLean R. Chapter 19 - Conclusions and recommendations. In: *The Vale of Leven Hospital Inquiry Report*. The Vale of Leven Hospital Inquiry, 2014. Available at: https://hub.careinspectorate.com/media/1415/vale-of-leven-hospital-inquiry-report.pdf. Accessed July 20, 2023.

425. Gebel J, Gemein S, Kampf G, *et al.* Isopropanol at 60% and at 70% are effective against 'isopropanol-tolerant' Enterococcus faecium. *J Hosp Infect* 2019; **103**: e88–91.

426. National Services Scotland. *Acute Hospital Activity and NHS Beds Information in Scotland, annual 2017 release*. National Services Scotland; 2018. Available at: https://www.isdscotland.org/Health-Topics/Hospital-Care/Publications/index.asp.

427. Liese J, Schüle L, Oberhettinger P, *et al.* Expansion of Vancomycin-resistant *Enterococcus faecium* in an academic tertiary hospital in Southwest Germany: a large-scale whole genome-based outbreak investigation. *Antimicrob Agents Chemother* 2019; **63**: e01978-18.

428. Yee R, Simner PJ. Next-Generation Sequencing Approaches to Predicting Antimicrobial Susceptibility Testing Results. *Clin Lab Med* 2022; **42**: 557–72.

429. Hunt M, Bradley P, Lapierre SG, *et al.* Antibiotic resistance prediction for Mycobacterium tuberculosis from genome sequence data with Mykrobe. *Wellcome Open Res* 2019; **4**: 191.

430. Anahtar MN, Bramante JT, Xu J, *et al.* Prediction of Antimicrobial Resistance in Clinical Enterococcus faecium Isolates Using a Rules-Based Analysis of Whole-Genome Sequences. *Antimicrob Agents Chemother* 2022; **66**: e01196-21.

431. Tyson GH, Sabo JL, Rice-Trujillo C, *et al.* Whole-genome sequencing based characterization of antimicrobial resistance in Enterococcus. *Pathog Dis* 2018; **76**: fty018.

432. Penven M, Zouari A, Nogues S, *et al.* Web-based prediction of antimicrobial resistance in enterococcal clinical isolates by whole-genome sequencing. *Eur J Clin Microbiol Infect Dis* 2023; **42**: 67–76.

433. Coll F, Gouliouris T, Blane B, *et al.* Improved Accuracy of Antibiotic Resistance Determination from *Enterococcus faecium* Whole-Genome Sequences. 2023. Available at: https://www.ssrn.com/abstract=4345808. Accessed May 25, 2023.

434. Sherry NL, Horan KA, Ballard SA, *et al.* An ISO-certified genomics workflow for identification and surveillance of antimicrobial resistance. *Nat Commun* 2023; **14**: 60.

435. Arredondo-Alonso S, Willems RJ, van Schaik W, *et al.* On the (im)possibility of reconstructing plasmids from whole-genome short-read sequencing data. *Microb Genomics* 2017; **3**: e000128.

436. Top J, Willems R, Bonten M. Emergence of CC17 Enterococcus faecium: From commensal to hospital-adapted pathogen. *FEMS Immunol Med Microbiol* 2008; **52**: 297–308.

437. Gouliouris T, Raven KE, Ludden C, *et al.* Genomic Surveillance of Enterococcus faecium Reveals Limited Sharing of Strains and Resistance Genes between Livestock and Humans in the United Kingdom. *mBio* 2018; **9**: e01780-18.

438. Coombs GW, Daley DA, Mowlaboccus S, *et al.* Australian Group on Antimicrobial Resistance (AGAR) Australian Enterococcal Sepsis Outcome Programme (AESOP) Annual Report 2019. *Commun Dis Intell* 2020; **44**. Available at: https://www1.health.gov.au/internet/main/publishing.nsf/Content/AD2DF748753AFDE1CA2584E2008009BA/$File/australian_group_on_antimicrobial_resistance_agar_australian_enterococcal_sepsis_outcome_programme_aesop_annual_report_2019.pdf. Accessed July 30, 2023.

439. Brenciani A, Morroni G, Schwarz S, *et al.* Oxazolidinones: mechanisms of resistance and mobile genetic elements involved. *J Antimicrob Chemother* 2022; **77**: 2596–621.

440. Shen W, Zhang R, Cai J. Co-occurrence of multiple plasmid-borne linezolid resistance genes-optrA, cfr, poxtA2 and cfr(D) in an Enterococcus faecalis isolate from retail meat. *J Antimicrob Chemother* 2023; **78**: 1637–43.

441. Bates J. Epidemiology of vancomycin-resistant enterococci in the community and the relevance of farm animals to human infection. *J Hosp Infect* 1997; **37**: 89–101.

442. Bates J, Jordens Z, Selkon JB. Evidence for an animal origin of vancomycin-resistant enterococci. *The Lancet* 1993; **342**: 490–1.

443. European Commission. *Ban on the antibiotic 'Avoparcin' in animal feed*. Brussels, Belgium: European Commission; 1997. Available at: https://ec.europa.eu/commission/presscorner/detail/en/IP_97_71. Accessed July 30, 2023.

444. Garcia-Migura L, Pleydell E, Barnes S, *et al.* Characterization of Vancomycin-Resistant Enterococcus faecium Isolates from Broiler Poultry and Pig Farms in England and Wales. *J Clin Microbiol* 2005; **43**: 3283–9.

445. Wick RR, Judd LM, Wyres KL, *et al.* Recovery of small plasmid sequences via Oxford Nanopore sequencing. *Microb Genomics* 2021; **7**: 000631.

446. Wick RR, Judd LM, Gorrie CL, *et al.* Completing bacterial genome assemblies with multiplex MinION sequencing. *Microb Genomics* 2017; **3**: e000132.

447. Foster-Nyarko E, Cottingham H, Wick RR, *et al.* Nanopore-only assemblies for genomic surveillance of the global priority drug-resistant pathogen, Klebsiella pneumoniae. *Microb Genomics* 2023; **9**: mgen000936.

448. Leigh RJ, McKenna C, McWade R, *et al.* Comparative genomics and pangenomics of vancomycin-resistant and susceptible Enterococcus faecium from Irish hospitals. *J Med Microbiol* 2022; **71**: 001590.

449. van Hal SJ, Ip CLC, Ansari MA, *et al.* Evolutionary dynamics of Enterococcus faecium reveals complex genomic relationships between isolates with independent emergence of vancomycin resistance. *Microb Genomics* 2016; **2**: e000048.

450. Hansen SGK, Roer L, Karstensen KT, *et al.* Vancomycin-sensitive Enterococcus faecium bacteraemia – hospital transmission and mortality in a Danish University Hospital. *J Med Microbiol* 2023; **72**: 001731.

451. Sundermann AJ, Babiker A, Marsh JW, *et al.* Outbreak of Vancomycin-resistant Enterococcus faecium in Interventional Radiology: Detection Through Whole Genome Sequencing-Based Surveillance. *Clin Infect Dis* 2020; **70**: 2336–43.

452. Munk P, Brinch C, Møller FD, *et al.* Genomic analysis of sewage from 101 countries reveals global landscape of antimicrobial resistance. *Nat Commun* 2022; **13**: 7251.

453. Kang X, Luo X, Schönhuth A. StrainXpress: strain aware metagenome assembly from short reads. *Nucleic Acids Res* 2022; **50**: e101.

Appendices

Appendix 1

Information on hybrid genome assemblies for *optrA*+ *E. faecalis*

Isolate	Year Isolated	ST	Plasmid	Copy Number[a]	Size (bp)	Plasmid rep type	Best NCBI match	Antimicrobial and Heavy Metal	Amino acid substitutions compared	Nucleotide substitutions compared	Transposases (n)
BX5936	2017	894	pBX5936-1	1	68656	rep9	Efs pE035; coverage 65%; 98% ID (MK140641)	*fexA; optrA*	S2F	C5T	ISEf1 (2); IS1216 (2)
			pBX5936-2	1	51669	rep9	Efs FC unnamed plasmid1; coverage 85%;	*catA8; tet(L); tet(M);*	-	-	ND

							100% ID (CP028836)	*ant(6)-Ia; cadA; copZ; ermB*			
BX8117	2017	16	pBX8117-1	1	68773	rep9	Efs FDAARGOS_324 unnamed plasmid2; coverage 100%; 100% ID (CP028284)	None	-	-	ND
			pBX8117-2	1	41839	rep9	Efs pEF123; coverage 64%; 98% ID (KX579977)	*catA8; cfr(D); optrA; fexA*	K3E; N12Y; E37K; N122K; Y135C;	A7G; A34T; G109A; G240A; C366A;	IS*1216* (5)

									Y176D; A350V; V395A; A396S; Q509K; Q541E; M552L; N560Y; K562N; Q565K; E614Q; I627L; D633E; N640I; R650G	A404G; C474T; T526G; A582T; C606T; T894C; G1035A; A1041G; C1049T; C1184C; G1186T; T1197C; T1449A; A1464G; A1491C; T1500C;	

										C1525A; T1530C; A1587G; C1621G; T1626C; A1654T; T1662C; A1678T; A1686T; C1693A; C1737T; G1840C; A1879T; T1899A; A1919T; A1948G	

TM6294	2017	585	pTM6294-1	1	75362	rep9	Efs FC unnamed plasmid1; coverage 74%; 99% ID (CP028836)	*catA8; tet(L); tet(M); ant(6)-Ia; cadA; copZ; ermB; aph(3')-IIIa; sat4; ant(6)-Ia; lnuB; lsaE; ant(9); ant(6)-Ia; aac(6')-Ie-*	-	-	ND

								aph(2'')-Ia; aadK; ermB; dfrG			
			pTM6294-2	1	52776	rep9	Efs pE035; coverage 87%; 99% ID (MK140641)	*fexA; optrA*	None	A1533T	ISL3-family (1); IS1216 (1)
WE0254	2015	19	pWE0254-1	1	80496	repUS11	Efs FDAARGOS_324 unnamed plasmid3; coverage 49%; 99% ID (CP028283)	*ant(9)-Ia; ermA-like; fexA; optrA*	None	G1779A; C1833T	IS3-family (8); IS1216-partial (1)

WE0438	2016	330	pWE0438	1	61284	rep9	Efs pEF123; coverage 76%; 99% ID (KX579977)	*tet(L); tet(M); bcrA; cadA; copZ ; ant(6)-Ia; optrA; fexA; ermB*	K3E; Y176D; I622M	A7G; T526G; C849T; A1866G	IS1216 (1); ISEnfa1 (2); IS3-family (4); IS6-family (1); Tn3-family (2)
WE0851	2014	480	pWE0851-1	1	59708	repUS11	Efs pEF123; coverage 22%; 100% ID (KX579977)	*fexA; optrA; ermA-like*	T112K; Y176D	C335A; T526G	IS1216 (1); IS3-family (6); Tn3-family (1)
			pWE0851-2	1	26996	repUS11	Efs pKUB3007-3; coverage 63%;	*aac(6')-Ie-*	-	-	ND

							100% ID (AP018546)	*aph(2'')-Ia*			
			pWE0851-3	3	10826	None detected	Efs pE035; coverage 63%; 99% ID (MK140641)	*aac(6')-Ie-aph(2'')-Ia; aac(6')-Ie-aph(2'')-Ia; aadK; ermB; ant(6)-Ia; aph(3')-IIIa; sat4*	-	-	ND

bp, base pairs; Efs, E. faecalis; ID, identity; NCBI, National Center for Biotechnology Information ; ND, not determined; ST, sequence type

a Inferred from depth of coverage relative to chromosomal contigs in hybrid assembly

b Amino acid sequence variants compared to the first described optrA sequence from pE394 (KP399637)

Appendix 2

Genome components identified in VRED06-02 and VRED06-10 hybrid assemblies

Isolate	Element	Size (bp)	CDS (n)	Circular	Copies	*rep* type	Prophages	Resistance genes	Virulence genes
VRED06-02 (ST1424)	Chromosome	2,945,113	2774	Yes	1	-	2 complete (43.2 kb, 52.3 kb), 1 questionable (21.5 kb), and 5 incomplete (6.7 kb, 14.3 kb, 16.6 kb, 24.7 kb, 27.5 kb)	*aac(6')-Ii, ant(9)-Ia, dfrG, erm(A), msr(C), tet(M)*	*acm, efaA, esp*
	p1_VRED06-02	201,362	207	Yes	2	US15	2 complete (42.6 kb, 31.3 kb), and 2 questionable (14.5 kb, 28.4 kb)	*aac(6')-aph(2'')*	hyl_{Efm}

	p2_VRED06-02	83,608	89	Yes	5	2, 18b	1 complete (34.4 kb)	*aph(3')-III, erm(B), vanA*	-
	p3_VRED06-02	59,532	67	Yes	2	17	-	-	-
	p4_VRED06-02	6,302	8	Yes	14	-	-	-	-
	p5_VRED06-02	5,212	6	Yes	9	18b	-	-	-
	p6_VRED06-02	2,947	4	Yes	12	14b	-	-	-
	p7_VRED06-02	4,372	6	Yes	14	14b	-	-	-
VRED06-10 (ST80)	Chromosome	2,814,943	2658	Yes	1	-	2 intact (49.2 kb, 53.6 kb), 2 questionable (6.2 kb, 14.8 kb), and 2 incomplete (7.2 kb, 27.5 kb)	*aac(6')-Ii, msrC*	*acm, efaA*

	p1_VRED06-10	205,024	208	Yes	1	1, US15	2 incomplete (5.1 kb, 28.8 kb)	*aac(6')-aph(2'')*, *erm(B)*	-
	p2_VRED06-10	150,852	183	No	1	-	1 questionable (16.6 kb), and 2 incomplete (14.6 kb, 16.6k b)	*vanA*	-
	p3_VRED06-10	51,924	61	Yes	2	2, 17	1 complete (24.9 kb)	*ant(6)-Ia*, *aph(3'')-III*, *erm(B)*, *tet(S)*	-
	p4_VRED06-10	6,173	7	Yes	9	11a	-	-	-
	p5_VRED06-10	4,464	3	Yes	6	-	-	-	-

bp, base pairs; CDS, coding sequence; n, number

Appendix 3

Residuals in Chi-square analysis of MLST to AMR, plasmid *rep* type, and virulence genes in national collection

Target				**ST203 (n=125)**	**ST80 (n=117)**	**Other STs (n=84)**
AMR	Aminoglycosides	Gentamicin, Kanamycin	*aac(6')-Ie-aph(2'')-Ia*	-2.511	2.596	-0.09
		Kanamycin	*aph (3')-IIIa* or *aac(6')-Ie-aph(2'')*	1.238	0.193	-1.792
		Streptomycin	*ant(6)-Ia*	2.397	0.215	-3.272
	Macrolides	Erythromycin	*ermB*	1.103	-0.313	-0.993
		Erythromycin	*ermT*	-4.093	0.595	4.387
	Tetracyclines	Tetracycline	*tet*(M)	6.68	-3.916	-3.489
		Tetracycline	*tet*(S)	-5.013	7.847	-3.501
		Tetracycline	*tet*(L)	-2.081	-3.683	7.199
	Diaminopyrimidines	Trimethoprim	*dfrG*	-5.95	-1.358	9.149
	Fluoroquinolones	Ciprofloxacin	*gyrA* S83Y	-5.815	-1.246	8.84

		Ciprofloxacin	*gyrA* S83I	3.992	0.334	-5.418
		Ciprofloxacin	*parC* S80I	-8.157	5.85	2.929
		Ciprofloxacin	*parC* S80R	9.208	-7.23	-2.525
Plasmids			*rep1*	-3.618	5.57	-2.727
			rep2	2.669	-0.564	-2.228
			rep11a	2.767	0.922	-4.111
			rep14a	-0.954	-2.046	3.484
			rep14b	-1.565	-1.596	3.61
			rep17	4.422	-1.532	-2.976
			rep18b	-2.44	-2.485	5.624
			rep29	-2.683	-2.612	6.041
			repUS7	-2.898	4.506	-2.237
			repUS12	-3.266	-1.06	4.818
			repUS43	-5.052	2.002	3.1
Virulence			*efaA*	17.245	18.115	1.8
			hyl	29.162	30.633	3.044

AMR, antimicrobial resistance; n, number

Appendix 4

Full PopPUNK clustering for Scottish and international genomes

VLKC	International (%) [n=1584]	Scottish (%) [n=326]	Total (%) [n=1910]
6_12_17_23_30	755 (47.7)	146 (44.8)	901 (47.2)
1_16_21	14 (0.9)	95 (29.1)	109 (5.7)
7	1 (0.1)	21 (6.4)	22 (1.2)
8	0 (0)	20 (6.1)	20 (1.1)
43	0 (0)	14 (4.3)	14 (0.7)
5	19 (1.2)	7 (2.2)	26 (1.4)
35	18 (1.1)	5 (1.5)	23 (1.2)
56	0 (0)	4 (1.2)	4 (0.2)
61	0 (0)	3 (0.9)	3 (0.2)
73	0 (0)	2 (0.6)	2 (0.1)
2	26 (1.6)	1 (0.3)	27 (1.4)
13	16 (1)	1 (0.3)	17 (0.9)
11	15 (1)	1 (0.3)	16 (0.8)
115	0 (0)	1 (0.3)	1 (0.1)
116	0 (0)	1 (0.3)	1 (0.1)
117	0 (0)	1 (0.3)	1 (0.1)
118	0 (0)	1 (0.3)	1 (0.1)
119	0 (0)	1 (0.3)	1 (0.1)

120	0 (0)	1 (0.3)	1 (0.1)
39	32 (2)	0 (0)	32 (1.7)
40	27 (1.7)	0 (0)	27 (1.4)
41	20 (1.3)	0 (0)	20 (1.1)
42	15 (1)	0 (0)	15 (0.8)
44	12 (0.8)	0 (0)	12 (0.6)
45	12 (0.8)	0 (0)	12 (0.6)
46	11 (0.7)	0 (0)	11 (0.6)
47	10 (0.6)	0 (0)	10 (0.5)
48	10 (0.6)	0 (0)	10 (0.5)
49	8 (0.5)	0 (0)	8 (0.4)
9	7 (0.4)	0 (0)	7 (0.4)
50	6 (0.4)	0 (0)	6 (0.3)
51	5 (0.3)	0 (0)	5 (0.3)
52	5 (0.3)	0 (0)	5 (0.3)
53	5 (0.3)	0 (0)	5 (0.3)
54	5 (0.3)	0 (0)	5 (0.3)
55	5 (0.3)	0 (0)	5 (0.3)
57	4 (0.3)	0 (0)	4 (0.2)
58	4 (0.3)	0 (0)	4 (0.2)
59	4 (0.3)	0 (0)	4 (0.2)
60	4 (0.3)	0 (0)	4 (0.2)
62	3 (0.2)	0 (0)	3 (0.2)

63	3 (0.2)	0 (0)	3 (0.2)
64	3 (0.2)	0 (0)	3 (0.2)
65	3 (0.2)	0 (0)	3 (0.2)
66	3 (0.2)	0 (0)	3 (0.2)
67	3 (0.2)	0 (0)	3 (0.2)
68	3 (0.2)	0 (0)	3 (0.2)
69	3 (0.2)	0 (0)	3 (0.2)
70	3 (0.2)	0 (0)	3 (0.2)
71	3 (0.2)	0 (0)	3 (0.2)
72	3 (0.2)	0 (0)	3 (0.2)
74	2 (0.1)	0 (0)	2 (0.1)
75	2 (0.1)	0 (0)	2 (0.1)
76	2 (0.1)	0 (0)	2 (0.1)
77	2 (0.1)	0 (0)	2 (0.1)
78	2 (0.1)	0 (0)	2 (0.1)
79	2 (0.1)	0 (0)	2 (0.1)
80	2 (0.1)	0 (0)	2 (0.1)
81	2 (0.1)	0 (0)	2 (0.1)
82	2 (0.1)	0 (0)	2 (0.1)
83	2 (0.1)	0 (0)	2 (0.1)
84	2 (0.1)	0 (0)	2 (0.1)
85	2 (0.1)	0 (0)	2 (0.1)
86	2 (0.1)	0 (0)	2 (0.1)

87	2 (0.1)	0 (0)	2 (0.1)
88	2 (0.1)	0 (0)	2 (0.1)
89	2 (0.1)	0 (0)	2 (0.1)
90	2 (0.1)	0 (0)	2 (0.1)
91	2 (0.1)	0 (0)	2 (0.1)
92	2 (0.1)	0 (0)	2 (0.1)
93	2 (0.1)	0 (0)	2 (0.1)
94	2 (0.1)	0 (0)	2 (0.1)
95	2 (0.1)	0 (0)	2 (0.1)
96	2 (0.1)	0 (0)	2 (0.1)
97	2 (0.1)	0 (0)	2 (0.1)
98	2 (0.1)	0 (0)	2 (0.1)
99	2 (0.1)	0 (0)	2 (0.1)
100	2 (0.1)	0 (0)	2 (0.1)
101	2 (0.1)	0 (0)	2 (0.1)
102	2 (0.1)	0 (0)	2 (0.1)
103	2 (0.1)	0 (0)	2 (0.1)
104	2 (0.1)	0 (0)	2 (0.1)
105	2 (0.1)	0 (0)	2 (0.1)
106	2 (0.1)	0 (0)	2 (0.1)
107	2 (0.1)	0 (0)	2 (0.1)
108	2 (0.1)	0 (0)	2 (0.1)
109	2 (0.1)	0 (0)	2 (0.1)

110	2 (0.1)	0 (0)	2 (0.1)
111	2 (0.1)	0 (0)	2 (0.1)
112	2 (0.1)	0 (0)	2 (0.1)
113	2 (0.1)	0 (0)	2 (0.1)
114	2 (0.1)	0 (0)	2 (0.1)
26_33	1 (0.1)	0 (0)	1 (0.1)
121	1 (0.1)	0 (0)	1 (0.1)
122	1 (0.1)	0 (0)	1 (0.1)
123	1 (0.1)	0 (0)	1 (0.1)
124	1 (0.1)	0 (0)	1 (0.1)
125	1 (0.1)	0 (0)	1 (0.1)
126	1 (0.1)	0 (0)	1 (0.1)
127	1 (0.1)	0 (0)	1 (0.1)
128	1 (0.1)	0 (0)	1 (0.1)
129	1 (0.1)	0 (0)	1 (0.1)
130	1 (0.1)	0 (0)	1 (0.1)
131	1 (0.1)	0 (0)	1 (0.1)
132	1 (0.1)	0 (0)	1 (0.1)
133	1 (0.1)	0 (0)	1 (0.1)
134	1 (0.1)	0 (0)	1 (0.1)
135	1 (0.1)	0 (0)	1 (0.1)
136	1 (0.1)	0 (0)	1 (0.1)
137	1 (0.1)	0 (0)	1 (0.1)

138	1 (0.1)	0 (0)	1 (0.1)
139	1 (0.1)	0 (0)	1 (0.1)
140	1 (0.1)	0 (0)	1 (0.1)
141	1 (0.1)	0 (0)	1 (0.1)
142	1 (0.1)	0 (0)	1 (0.1)
143	1 (0.1)	0 (0)	1 (0.1)
144	1 (0.1)	0 (0)	1 (0.1)
145	1 (0.1)	0 (0)	1 (0.1)
146	1 (0.1)	0 (0)	1 (0.1)
147	1 (0.1)	0 (0)	1 (0.1)
148	1 (0.1)	0 (0)	1 (0.1)
149	1 (0.1)	0 (0)	1 (0.1)
150	1 (0.1)	0 (0)	1 (0.1)
151	1 (0.1)	0 (0)	1 (0.1)
152	1 (0.1)	0 (0)	1 (0.1)
153	1 (0.1)	0 (0)	1 (0.1)
154	1 (0.1)	0 (0)	1 (0.1)
155	1 (0.1)	0 (0)	1 (0.1)
156	1 (0.1)	0 (0)	1 (0.1)
157	1 (0.1)	0 (0)	1 (0.1)
158	1 (0.1)	0 (0)	1 (0.1)
159	1 (0.1)	0 (0)	1 (0.1)
160	1 (0.1)	0 (0)	1 (0.1)

161	1 (0.1)	0 (0)	1 (0.1)
162	1 (0.1)	0 (0)	1 (0.1)
163	1 (0.1)	0 (0)	1 (0.1)
164	1 (0.1)	0 (0)	1 (0.1)
165	1 (0.1)	0 (0)	1 (0.1)
166	1 (0.1)	0 (0)	1 (0.1)
167	1 (0.1)	0 (0)	1 (0.1)
168	1 (0.1)	0 (0)	1 (0.1)
169	1 (0.1)	0 (0)	1 (0.1)
170	1 (0.1)	0 (0)	1 (0.1)
171	1 (0.1)	0 (0)	1 (0.1)
172	1 (0.1)	0 (0)	1 (0.1)
173	1 (0.1)	0 (0)	1 (0.1)
174	1 (0.1)	0 (0)	1 (0.1)
175	1 (0.1)	0 (0)	1 (0.1)
176	1 (0.1)	0 (0)	1 (0.1)
177	1 (0.1)	0 (0)	1 (0.1)
178	1 (0.1)	0 (0)	1 (0.1)
179	1 (0.1)	0 (0)	1 (0.1)
180	1 (0.1)	0 (0)	1 (0.1)
181	1 (0.1)	0 (0)	1 (0.1)
182	1 (0.1)	0 (0)	1 (0.1)
183	1 (0.1)	0 (0)	1 (0.1)

184	1 (0.1)	0 (0)	1 (0.1)
185	1 (0.1)	0 (0)	1 (0.1)
186	1 (0.1)	0 (0)	1 (0.1)
187	1 (0.1)	0 (0)	1 (0.1)
188	1 (0.1)	0 (0)	1 (0.1)
189	1 (0.1)	0 (0)	1 (0.1)
190	1 (0.1)	0 (0)	1 (0.1)
191	1 (0.1)	0 (0)	1 (0.1)
192	1 (0.1)	0 (0)	1 (0.1)
193	1 (0.1)	0 (0)	1 (0.1)
194	1 (0.1)	0 (0)	1 (0.1)
195	1 (0.1)	0 (0)	1 (0.1)
196	1 (0.1)	0 (0)	1 (0.1)
197	1 (0.1)	0 (0)	1 (0.1)
198	1 (0.1)	0 (0)	1 (0.1)
199	1 (0.1)	0 (0)	1 (0.1)
200	1 (0.1)	0 (0)	1 (0.1)
201	1 (0.1)	0 (0)	1 (0.1)
202	1 (0.1)	0 (0)	1 (0.1)
203	1 (0.1)	0 (0)	1 (0.1)
204	1 (0.1)	0 (0)	1 (0.1)
205	1 (0.1)	0 (0)	1 (0.1)
206	1 (0.1)	0 (0)	1 (0.1)

207	1 (0.1)	0 (0)	1 (0.1)
208	1 (0.1)	0 (0)	1 (0.1)
209	1 (0.1)	0 (0)	1 (0.1)
210	1 (0.1)	0 (0)	1 (0.1)
211	1 (0.1)	0 (0)	1 (0.1)
212	1 (0.1)	0 (0)	1 (0.1)
213	1 (0.1)	0 (0)	1 (0.1)
214	1 (0.1)	0 (0)	1 (0.1)
215	1 (0.1)	0 (0)	1 (0.1)
216	1 (0.1)	0 (0)	1 (0.1)
217	1 (0.1)	0 (0)	1 (0.1)
218	1 (0.1)	0 (0)	1 (0.1)
219	1 (0.1)	0 (0)	1 (0.1)
220	1 (0.1)	0 (0)	1 (0.1)
221	1 (0.1)	0 (0)	1 (0.1)
222	1 (0.1)	0 (0)	1 (0.1)
223	1 (0.1)	0 (0)	1 (0.1)
224	1 (0.1)	0 (0)	1 (0.1)
225	1 (0.1)	0 (0)	1 (0.1)
226	1 (0.1)	0 (0)	1 (0.1)
227	1 (0.1)	0 (0)	1 (0.1)
228	1 (0.1)	0 (0)	1 (0.1)
229	1 (0.1)	0 (0)	1 (0.1)

230	1 (0.1)	0 (0)	1 (0.1)
231	1 (0.1)	0 (0)	1 (0.1)
232	1 (0.1)	0 (0)	1 (0.1)
233	1 (0.1)	0 (0)	1 (0.1)
234	1 (0.1)	0 (0)	1 (0.1)
235	1 (0.1)	0 (0)	1 (0.1)
236	1 (0.1)	0 (0)	1 (0.1)
237	1 (0.1)	0 (0)	1 (0.1)
238	1 (0.1)	0 (0)	1 (0.1)
239	1 (0.1)	0 (0)	1 (0.1)
240	1 (0.1)	0 (0)	1 (0.1)
241	1 (0.1)	0 (0)	1 (0.1)
242	1 (0.1)	0 (0)	1 (0.1)
243	1 (0.1)	0 (0)	1 (0.1)
244	1 (0.1)	0 (0)	1 (0.1)
245	1 (0.1)	0 (0)	1 (0.1)
246	1 (0.1)	0 (0)	1 (0.1)
247	1 (0.1)	0 (0)	1 (0.1)
248	1 (0.1)	0 (0)	1 (0.1)
249	1 (0.1)	0 (0)	1 (0.1)
250	1 (0.1)	0 (0)	1 (0.1)
251	1 (0.1)	0 (0)	1 (0.1)
252	1 (0.1)	0 (0)	1 (0.1)

253	1 (0.1)	0 (0)	1 (0.1)
254	1 (0.1)	0 (0)	1 (0.1)
255	1 (0.1)	0 (0)	1 (0.1)
256	1 (0.1)	0 (0)	1 (0.1)
257	1 (0.1)	0 (0)	1 (0.1)
258	1 (0.1)	0 (0)	1 (0.1)
259	1 (0.1)	0 (0)	1 (0.1)
260	1 (0.1)	0 (0)	1 (0.1)
261	1 (0.1)	0 (0)	1 (0.1)
262	1 (0.1)	0 (0)	1 (0.1)
263	1 (0.1)	0 (0)	1 (0.1)
264	1 (0.1)	0 (0)	1 (0.1)
265	1 (0.1)	0 (0)	1 (0.1)
266	1 (0.1)	0 (0)	1 (0.1)
267	1 (0.1)	0 (0)	1 (0.1)
268	1 (0.1)	0 (0)	1 (0.1)
269	1 (0.1)	0 (0)	1 (0.1)
270	1 (0.1)	0 (0)	1 (0.1)
271	1 (0.1)	0 (0)	1 (0.1)
272	1 (0.1)	0 (0)	1 (0.1)
273	1 (0.1)	0 (0)	1 (0.1)
274	1 (0.1)	0 (0)	1 (0.1)
275	1 (0.1)	0 (0)	1 (0.1)

276	1 (0.1)	0 (0)	1 (0.1)
277	1 (0.1)	0 (0)	1 (0.1)
278	1 (0.1)	0 (0)	1 (0.1)
279	1 (0.1)	0 (0)	1 (0.1)
280	1 (0.1)	0 (0)	1 (0.1)
281	1 (0.1)	0 (0)	1 (0.1)
282	1 (0.1)	0 (0)	1 (0.1)
283	1 (0.1)	0 (0)	1 (0.1)
284	1 (0.1)	0 (0)	1 (0.1)
285	1 (0.1)	0 (0)	1 (0.1)
286	1 (0.1)	0 (0)	1 (0.1)
287	1 (0.1)	0 (0)	1 (0.1)
288	1 (0.1)	0 (0)	1 (0.1)
289	1 (0.1)	0 (0)	1 (0.1)
290	1 (0.1)	0 (0)	1 (0.1)
291	1 (0.1)	0 (0)	1 (0.1)
292	1 (0.1)	0 (0)	1 (0.1)
293	1 (0.1)	0 (0)	1 (0.1)
294	1 (0.1)	0 (0)	1 (0.1)
295	1 (0.1)	0 (0)	1 (0.1)
296	1 (0.1)	0 (0)	1 (0.1)
297	1 (0.1)	0 (0)	1 (0.1)
298	1 (0.1)	0 (0)	1 (0.1)

299	1 (0.1)	0 (0)	1 (0.1)
300	1 (0.1)	0 (0)	1 (0.1)
301	1 (0.1)	0 (0)	1 (0.1)
302	1 (0.1)	0 (0)	1 (0.1)
303	1 (0.1)	0 (0)	1 (0.1)
304	1 (0.1)	0 (0)	1 (0.1)
305	1 (0.1)	0 (0)	1 (0.1)
306	1 (0.1)	0 (0)	1 (0.1)
307	1 (0.1)	0 (0)	1 (0.1)
308	1 (0.1)	0 (0)	1 (0.1)
309	1 (0.1)	0 (0)	1 (0.1)
310	1 (0.1)	0 (0)	1 (0.1)
311	1 (0.1)	0 (0)	1 (0.1)
312	1 (0.1)	0 (0)	1 (0.1)
313	1 (0.1)	0 (0)	1 (0.1)
314	1 (0.1)	0 (0)	1 (0.1)
315	1 (0.1)	0 (0)	1 (0.1)
316	1 (0.1)	0 (0)	1 (0.1)
317	1 (0.1)	0 (0)	1 (0.1)
318	1 (0.1)	0 (0)	1 (0.1)
319	1 (0.1)	0 (0)	1 (0.1)
320	1 (0.1)	0 (0)	1 (0.1)
321	1 (0.1)	0 (0)	1 (0.1)

322	1 (0.1)	0 (0)	1 (0.1)
323	1 (0.1)	0 (0)	1 (0.1)
324	1 (0.1)	0 (0)	1 (0.1)
325	1 (0.1)	0 (0)	1 (0.1)
326	1 (0.1)	0 (0)	1 (0.1)
327	1 (0.1)	0 (0)	1 (0.1)
328	1 (0.1)	0 (0)	1 (0.1)
329	1 (0.1)	0 (0)	1 (0.1)
330	1 (0.1)	0 (0)	1 (0.1)
331	1 (0.1)	0 (0)	1 (0.1)
332	1 (0.1)	0 (0)	1 (0.1)
333	1 (0.1)	0 (0)	1 (0.1)
334	1 (0.1)	0 (0)	1 (0.1)
335	1 (0.1)	0 (0)	1 (0.1)
336	1 (0.1)	0 (0)	1 (0.1)
337	1 (0.1)	0 (0)	1 (0.1)
338	1 (0.1)	0 (0)	1 (0.1)
339	1 (0.1)	0 (0)	1 (0.1)
340	1 (0.1)	0 (0)	1 (0.1)
341	1 (0.1)	0 (0)	1 (0.1)
342	1 (0.1)	0 (0)	1 (0.1)
343	1 (0.1)	0 (0)	1 (0.1)
344	1 (0.1)	0 (0)	1 (0.1)

345	1 (0.1)	0 (0)	1 (0.1)
346	1 (0.1)	0 (0)	1 (0.1)
347	1 (0.1)	0 (0)	1 (0.1)
348	1 (0.1)	0 (0)	1 (0.1)
349	1 (0.1)	0 (0)	1 (0.1)
350	1 (0.1)	0 (0)	1 (0.1)
351	1 (0.1)	0 (0)	1 (0.1)
352	1 (0.1)	0 (0)	1 (0.1)
353	1 (0.1)	0 (0)	1 (0.1)
354	1 (0.1)	0 (0)	1 (0.1)
355	1 (0.1)	0 (0)	1 (0.1)
356	1 (0.1)	0 (0)	1 (0.1)
357	1 (0.1)	0 (0)	1 (0.1)
358	1 (0.1)	0 (0)	1 (0.1)
359	1 (0.1)	0 (0)	1 (0.1)
360	1 (0.1)	0 (0)	1 (0.1)
361	1 (0.1)	0 (0)	1 (0.1)
362	1 (0.1)	0 (0)	1 (0.1)
363	1 (0.1)	0 (0)	1 (0.1)
364	1 (0.1)	0 (0)	1 (0.1)
365	1 (0.1)	0 (0)	1 (0.1)
366	1 (0.1)	0 (0)	1 (0.1)
367	1 (0.1)	0 (0)	1 (0.1)

368	1 (0.1)	0 (0)	1 (0.1)
369	1 (0.1)	0 (0)	1 (0.1)
370	1 (0.1)	0 (0)	1 (0.1)
371	1 (0.1)	0 (0)	1 (0.1)
372	1 (0.1)	0 (0)	1 (0.1)
373	1 (0.1)	0 (0)	1 (0.1)
374	1 (0.1)	0 (0)	1 (0.1)
375	1 (0.1)	0 (0)	1 (0.1)
376	1 (0.1)	0 (0)	1 (0.1)
377	1 (0.1)	0 (0)	1 (0.1)
378	1 (0.1)	0 (0)	1 (0.1)
379	1 (0.1)	0 (0)	1 (0.1)
380	1 (0.1)	0 (0)	1 (0.1)
381	1 (0.1)	0 (0)	1 (0.1)
382	1 (0.1)	0 (0)	1 (0.1)
383	1 (0.1)	0 (0)	1 (0.1)
384	1 (0.1)	0 (0)	1 (0.1)
385	1 (0.1)	0 (0)	1 (0.1)
386	1 (0.1)	0 (0)	1 (0.1)
387	1 (0.1)	0 (0)	1 (0.1)
388	1 (0.1)	0 (0)	1 (0.1)
389	1 (0.1)	0 (0)	1 (0.1)
390	1 (0.1)	0 (0)	1 (0.1)

391	1 (0.1)	0 (0)	1 (0.1)
392	1 (0.1)	0 (0)	1 (0.1)
393	1 (0.1)	0 (0)	1 (0.1)
394	1 (0.1)	0 (0)	1 (0.1)
395	1 (0.1)	0 (0)	1 (0.1)
396	1 (0.1)	0 (0)	1 (0.1)
397	1 (0.1)	0 (0)	1 (0.1)
398	1 (0.1)	0 (0)	1 (0.1)
399	1 (0.1)	0 (0)	1 (0.1)
400	1 (0.1)	0 (0)	1 (0.1)
401	1 (0.1)	0 (0)	1 (0.1)
402	1 (0.1)	0 (0)	1 (0.1)
403	1 (0.1)	0 (0)	1 (0.1)
404	1 (0.1)	0 (0)	1 (0.1)
405	1 (0.1)	0 (0)	1 (0.1)
406	1 (0.1)	0 (0)	1 (0.1)
407	1 (0.1)	0 (0)	1 (0.1)
408	1 (0.1)	0 (0)	1 (0.1)
409	1 (0.1)	0 (0)	1 (0.1)
410	1 (0.1)	0 (0)	1 (0.1)
411	1 (0.1)	0 (0)	1 (0.1)
412	1 (0.1)	0 (0)	1 (0.1)
413	1 (0.1)	0 (0)	1 (0.1)

414	1 (0.1)	0 (0)	1 (0.1)
415	1 (0.1)	0 (0)	1 (0.1)
416	1 (0.1)	0 (0)	1 (0.1)
417	1 (0.1)	0 (0)	1 (0.1)
418	1 (0.1)	0 (0)	1 (0.1)
419	1 (0.1)	0 (0)	1 (0.1)
420	1 (0.1)	0 (0)	1 (0.1)
421	1 (0.1)	0 (0)	1 (0.1)
422	1 (0.1)	0 (0)	1 (0.1)
423	1 (0.1)	0 (0)	1 (0.1)
424	1 (0.1)	0 (0)	1 (0.1)
425	1 (0.1)	0 (0)	1 (0.1)
426	1 (0.1)	0 (0)	1 (0.1)
427	1 (0.1)	0 (0)	1 (0.1)
428	1 (0.1)	0 (0)	1 (0.1)
429	1 (0.1)	0 (0)	1 (0.1)
430	1 (0.1)	0 (0)	1 (0.1)
431	1 (0.1)	0 (0)	1 (0.1)
432	1 (0.1)	0 (0)	1 (0.1)
433	1 (0.1)	0 (0)	1 (0.1)
434	1 (0.1)	0 (0)	1 (0.1)
435	1 (0.1)	0 (0)	1 (0.1)
436	1 (0.1)	0 (0)	1 (0.1)

437	1 (0.1)	0 (0)	1 (0.1)
438	1 (0.1)	0 (0)	1 (0.1)
439	1 (0.1)	0 (0)	1 (0.1)
440	1 (0.1)	0 (0)	1 (0.1)
441	1 (0.1)	0 (0)	1 (0.1)
442	1 (0.1)	0 (0)	1 (0.1)
443	1 (0.1)	0 (0)	1 (0.1)
444	1 (0.1)	0 (0)	1 (0.1)
445	1 (0.1)	0 (0)	1 (0.1)
446	1 (0.1)	0 (0)	1 (0.1)
447	1 (0.1)	0 (0)	1 (0.1)
448	1 (0.1)	0 (0)	1 (0.1)
449	1 (0.1)	0 (0)	1 (0.1)
450	1 (0.1)	0 (0)	1 (0.1)
451	1 (0.1)	0 (0)	1 (0.1)
452	1 (0.1)	0 (0)	1 (0.1)
453	1 (0.1)	0 (0)	1 (0.1)
454	1 (0.1)	0 (0)	1 (0.1)
455	1 (0.1)	0 (0)	1 (0.1)
456	1 (0.1)	0 (0)	1 (0.1)
457	1 (0.1)	0 (0)	1 (0.1)
458	1 (0.1)	0 (0)	1 (0.1)
459	1 (0.1)	0 (0)	1 (0.1)

460	1 (0.1)	0 (0)	1 (0.1)
461	1 (0.1)	0 (0)	1 (0.1)
462	1 (0.1)	0 (0)	1 (0.1)
463	1 (0.1)	0 (0)	1 (0.1)
464	1 (0.1)	0 (0)	1 (0.1)
465	1 (0.1)	0 (0)	1 (0.1)
466	1 (0.1)	0 (0)	1 (0.1)
467	1 (0.1)	0 (0)	1 (0.1)
468	1 (0.1)	0 (0)	1 (0.1)
469	1 (0.1)	0 (0)	1 (0.1)
470	1 (0.1)	0 (0)	1 (0.1)
471	1 (0.1)	0 (0)	1 (0.1)
472	1 (0.1)	0 (0)	1 (0.1)
473	1 (0.1)	0 (0)	1 (0.1)
474	1 (0.1)	0 (0)	1 (0.1)
475	1 (0.1)	0 (0)	1 (0.1)
476	1 (0.1)	0 (0)	1 (0.1)
477	1 (0.1)	0 (0)	1 (0.1)
478	1 (0.1)	0 (0)	1 (0.1)
479	1 (0.1)	0 (0)	1 (0.1)
480	1 (0.1)	0 (0)	1 (0.1)
481	1 (0.1)	0 (0)	1 (0.1)
482	1 (0.1)	0 (0)	1 (0.1)

483	1 (0.1)	0 (0)	1 (0.1)
484	1 (0.1)	0 (0)	1 (0.1)
485	1 (0.1)	0 (0)	1 (0.1)
486	1 (0.1)	0 (0)	1 (0.1)
487	1 (0.1)	0 (0)	1 (0.1)
488	1 (0.1)	0 (0)	1 (0.1)
489	1 (0.1)	0 (0)	1 (0.1)
490	1 (0.1)	0 (0)	1 (0.1)
491	1 (0.1)	0 (0)	1 (0.1)
492	1 (0.1)	0 (0)	1 (0.1)
493	1 (0.1)	0 (0)	1 (0.1)
494	1 (0.1)	0 (0)	1 (0.1)
495	1 (0.1)	0 (0)	1 (0.1)
496	1 (0.1)	0 (0)	1 (0.1)
497	1 (0.1)	0 (0)	1 (0.1)
498	1 (0.1)	0 (0)	1 (0.1)
499	1 (0.1)	0 (0)	1 (0.1)
500	1 (0.1)	0 (0)	1 (0.1)
501	1 (0.1)	0 (0)	1 (0.1)
502	1 (0.1)	0 (0)	1 (0.1)
503	1 (0.1)	0 (0)	1 (0.1)
504	1 (0.1)	0 (0)	1 (0.1)
505	1 (0.1)	0 (0)	1 (0.1)

506	1 (0.1)	0 (0)	1 (0.1)
507	1 (0.1)	0 (0)	1 (0.1)
508	1 (0.1)	0 (0)	1 (0.1)
509	1 (0.1)	0 (0)	1 (0.1)
510	1 (0.1)	0 (0)	1 (0.1)
511	1 (0.1)	0 (0)	1 (0.1)
512	1 (0.1)	0 (0)	1 (0.1)
513	1 (0.1)	0 (0)	1 (0.1)

Publications

MICROBIOLOGY

SHORT COMMUNICATION
McHugh *et al.*, *Microbiology* 2022;168:001137
DOI 10.1099/mic.0.001137

MICROBIOLOGY SOCIETY
OPEN ACCESS

Presence of *optrA*-mediated linezolid resistance in multiple lineages and plasmids of *Enterococcus faecalis* revealed by long read sequencing

Martin P. McHugh[1,2], Benjamin J. Parcell[1,3]†, Kerry A. Pettigrew[1]‡, Geoff Toner[2], Elham Khatamzas[2]§, Noha el Sakka[3], Anne Marie Karcher[3]†, Joanna Walker[3], Robert Weir[4], Danièle Meunier[5], Katie L. Hopkins[5], Neil Woodford[5], Kate E. Templeton[2], Stephen H. Gillespie[1] and Matthew T. G. Holden[1,*]

Abstract

Transferable linezolid resistance due to *optrA*, *poxtA*, *cfr* and *cfr*-like genes is increasingly detected in enterococci associated with animals and humans globally. We aimed to characterize the genetic environment of *optrA* in linezolid-resistant *Enterococcus faecalis* isolates from Scotland. Six linezolid-resistant *E. faecalis* isolated from urogenital samples were confirmed to carry the *optrA* gene by PCR. Short read (Illumina) sequencing showed the isolates were genetically distinct (>13900 core SNPs) and belonged to different MLST sequence types. Plasmid contents were examined using hybrid assembly of short and long read (Oxford Nanopore MinION) sequencing technologies. The *optrA* gene was located on distinct plasmids in each isolate, suggesting that transfer of a single plasmid did not contribute to *optrA* dissemination in this collection. pTM6294-2, BX5936-1 and pWE0438-1 were similar to *optrA*-positive plasmids from China and Japan, while the remaining three plasmids had limited similarity to other published examples. We identified the novel Tn*6993* transposon in pWE0254-1 carrying linezolid (*optrA*), macrolide (*ermB*) and spectinomycin [ANT(9)-Ia] resistance genes. OptrA amino acid sequences differed by 0–20 residues. We report multiple variants of *optrA* on distinct plasmids in diverse strains of *E. faecalis*. It is important to identify the selection pressures driving the emergence and maintenance of resistance against linezolid to retain the clinical utility of this antibiotic.

INTRODUCTION

Enterococcus faecalis and *Enterococcus faecium* are carried in the intestinal tract and are important opportunistic pathogens in humans [1]. Treatment of enterococcal infections is challenging due to intrinsic or acquired resistance to multiple antimicrobials including aminoglycosides, benzylpenicillin, cephalosporins, fluoroquinolones, macrolides, tetracyclines and trimethoprim. Among the remaining treatment options, clinical *E. faecium* isolates are usually resistant to amoxicillin, and resistance to vancomycin is increasingly common [2]. In contrast, *E. faecalis* typically remains susceptible to amoxicillin and resistance to vancomycin is uncommon. Where vancomycin cannot be used, treatment options against severe enterococcal infections are largely limited to daptomycin, linezolid or combination therapy and are further complicated by issues with efficacy, susceptibility or toxicity [1].

Oxazolidinones such as linezolid block protein synthesis by binding to the 50S ribosomal subunit and inhibit formation of the initiation complex [3]. Linezolid resistance is reported in ≤1% of bloodstream enterococcal isolates in the UK and is an important

Received 17 August 2021; Accepted 10 January 2022; Published 07 February 2022
Author affiliations: [1]School of Medicine, University of St Andrews, St Andrews, UK; [2]NHS Lothian Infection Service, Royal Infirmary of Edinburgh, Edinburgh, UK; [3]Medical Microbiology, Aberdeen Royal Infirmary, Aberdeen, UK; [4]Medical Microbiology, Forth Valley Royal Hospital, Larbert, UK; [5]Antimicrobial Resistance and Healthcare Associated Infections (AMRHAI) Reference Unit, National Infection Service, Public Health England, London, UK.
***Correspondence:** Matthew T. G. Holden, mtgh@st-andrews.ac.uk
Keywords: antimicrobial resistance; *Enterococcus faecalis*; linezolid; *optrA*; plasmid; Tn*6993*.
Abbreviations: SNP, single nucleotide polymorphism.
†Present address: Medical Microbiology, Ninewells Hospital, Dundee, UK
‡Present address: School of Social and Behavioural Sciences, Erasmus University, Rotterdam, Netherlands
§Present address: Department of Medicine III, University Hospital, LMU Munich, Germany.
Repositories: sequencing reads and annotated assemblies for this study are deposited in the European Nucleotide Archive at EMBL-EBI under accession number PRJEB36950 (https://www.ebi.ac.uk/ena/data/view/PRJEB36950).
One supplementary table and two supplementary figures are available with the online version of this article.

antimicrobial for the treatment of multi-drug-resistant Gram-positive infections, including vancomycin-resistant enterococci [4, 5]. The G2576T mutation in the chromosomal 23S rRNA genes can arise *de novo* during extended linezolid therapy [6], although antimicrobial stewardship and infection prevention and control measures appear to be successful in limiting the generation and spread of mutational linezolid resistance in clinical practice [7]. The methyltransferases Cfr, Cfr(B) and Cfr(D), and the ABC-F ribosomal protection proteins OptrA and PoxtA also confer resistance to linezolid in enterococci but are carried on mobile genetic elements, which can spread across genetically distinct lineages in the absence of antimicrobial selection [8–14]. Recent international surveillance confirmed that linezolid resistance remains rare, but *optrA* has recently spread to every continent and is the dominant mechanism of linezolid resistance in *E. faecalis* [15]. Surveillance has also detected *optrA* in the UK [16]. Studies into the genetic context of *optrA* have identified the gene on both the chromosome and plasmids, often associated with insertion sequences such as IS*1216*, a possible vehicle for the rapid spread of *optrA* [17, 18].

We used whole genome sequencing to determine whether Scottish *optrA*-positive *E. faecalis* isolates represent transmission of a single clonal lineage. We hypothesized that spread of *optrA* is driven by a single mobile genetic element, and to investigate this we made hybrid assemblies of short and long read sequencing data to generate complete genomes and to reconstruct the genetic environment of *optrA*.

METHODS

Bacterial strains

Study isolates were a convenience sample from three regional hospital laboratories during 2014–17; as such they may not reflect the entire Scottish population of *optrA*-positive *E. faecalis. E. faecalis* were identified from clinical samples using MALDI-TOF MS or the Vitek-2 GP-ID card (bioMérieux). Initial antimicrobial susceptibility testing was performed with the Vitek-2 AST-607 card; where linezolid resistance was detected the full MIC was determined by agar dilution methodology at the AMRHAI reference laboratory, and susceptibility testing was interpreted with EUCAST breakpoints [19]. Linezolid-resistant isolates were then screened for the genetic determinant of resistance at AMRHAI. Detection of the G2576T mutation (*Escherichia coli* numbering) in the 23S rRNA genes was investigated by PCR-RFLP or by a real-time PCR-based allelic discrimination assay [20, 21]. The *cfr* and *optrA* genes were sought by a multiplex PCR using primers for the detection of *cfr* (*cfr-fw*: 5′-TGAAGTATAAAGCAGGTTGG GAGTCA-3′ and *cfr-rev*: 5′-ACCATATAATTGACCACAAGCAGC-3′) [22] and *optrA* (*optrA-F*: 5′-GACCGGTGTCCTCTTT-GTCA-3′ and *optrA-R*: 5′-TCAATGGAGTTACGATCGCCT-3′) (AMRHAI, unpublished).

Access to isolates and clinical data was approved by the NHS Scotland Biorepository Network (Ref. TR000126).

Whole genome sequencing and genomic analysis

Genomic DNA was extracted from pelleted overnight broth cultures using the MasterPure Gram Positive DNA Purification Kit (Cambio), or QiaSymphony DSP DNA Mini Kit (Qiagen). Short read barcoded libraries were prepared using the Nextera XT kit (Illumina) and sequenced with a MiSeq instrument (Illumina) using 250 bp paired-end reads on a 500-cycle v2 kit. Short reads were quality trimmed with Trimmomatic v0.36 and the settings [LEADING:5 TRAILING:5 SLIDINGWINDOW:4:15 MINLEN:100] [23]. Barcoded long read libraries were generated with the 1D Ligation Sequencing Kit (Oxford Nanopore Technologies) and sequenced with an R9.4 flow cell on a MinION sequencer (Oxford Nanopore Technologies). Base-calling and barcode de-multiplexing was performed with Albacore v2.1.3 (Oxford Nanopore Technologies) and the resulting fast5 files were converted to fastq with Poretools v0.6.0 [24], or basecalled and de-multiplexed with Albacore v2.3.3 with direct fastq output. Porechop v0.2.3 (https://github.com/rrwick/Porechop) was used to remove chimeric reads and trim adapter sequences. Sequencing reads and annotated assemblies for this study have been deposited in the European Nucleotide Archive at EMBL-EBI under accession number PRJEB36950 (https://www.ebi.ac.uk/ena/data/view/PRJEB36950).

Short reads were mapped to the *E. faecalis* reference genome V583 (accession number AE016830) using SMALT v0.7.4 [25]. Mapped assemblies were aligned, and regions annotated as mobile genetic elements in the V583 genome (transposons, integrases, plasmids, phages, insertion sequences, resolvases and recombinases) were removed from the assembly (https://github.com/sanger-pathogens/remove_blocks_from_aln). All sites in the alignment with SNPs were extracted using SNP-sites v2.4.0 [26] and pairwise SNP counts were calculated (https://github.com/simonrharris/pairwise_difference_count).

MLST profiling was performed using SRST2 v0.2.0 [27] and the *E. faecalis* MLST database (https://pubmlst.org/efaecalis/) sited at the University of Oxford [28, 29]. Antimicrobial resistance mechanisms were detected using ARIBA v2.12.1 [30] and the ResFinder database v3.0 [31] with the addition of linezolid resistance mutations in the 23S rRNA (G2505A and G2576T based on *E. coli* numbering) and *rplC*, *rplD*, and *rplV* ribosomal protein genes.

Hybrid assembly was performed with Illumina short reads and Nanopore long reads using Unicycler v0.4.7 [32] in standard mode. The resulting assemblies were annotated with Prokka v1.5.1 using a genus-specific RefSeq database [33]. Hybrid assemblies were checked for indel errors using Ideel (https://github.com/mw55309/ideel) and UniProtKB TrEMBL database v2019_1. Plasmid comparisons were generated and visualized with EasyFig v2.2.2 [34].

RESULTS

Detection of *optrA* in distinct *E. faecalis* strains

There were 14133 isolates of *E. faecalis* during the study period from all sample types: 14 (0.1%) were identified as linezolid-resistant, and eight (57.1%) were confirmed as *optrA*-positive at the AMRHAI reference laboratory. Six *optrA*-positive *E. faecalis* were available for further characterization (Table 1). The earliest isolates in this collection were from the Grampian region in the northeast of Scotland in 2014, 2015 and 2016. Three more isolates were identified in 2017 from the Lothian and Forth Valley regions in east and central Scotland (Table 1), with no clear epidemiological links between the patients. Only one patient had known exposure to linezolid prior to the isolation of an *optrA*-positive *E. faecalis*, two patients were hospitalized at the time of sample collection while the remaining four were from general practice. Samples were collected for symptomatic urinary tract infection or orchitis.

Whole genome sequencing was performed to investigate the genetic relationship between the isolates. *In silico* MLST showed the six isolates belonged to different STs, suggesting they were genetically distinct (Table 1). To further confirm this, we analysed SNPs in the core genomes of the *optrA*-positive isolates and found the isolates differed by a median 18806 SNPs (range 13909–22272). Previous estimates suggest a genetic diversification rate of 2.5–3.4 SNPs/year for *E. faecalis*, highlighting the *optrA*-positive isolates share a very distant common ancestor [35].

optrA is carried on diverse genetic platforms

Hybrid assembly produced complete or near-complete genomes with <3% putative coding sequences shorter than the closest reference match. This indicated the hybrid assembly process removed most indel errors, with 1–5% of coding sequences expected to represent true truncated pseudogenes [36]. The hybrid assemblies contained between one and three plasmids ranging in size from 11 to 80 kb, with *optrA* present on a single complete plasmid in each isolate (pBX5936-1, pBX8117-2, pTM6294-2, pWE0254-1, pWE0438, pWE0851-1; Table S1, available in the online version of this article).

The *optrA*-positive plasmids shared limited sequence similarity to the first described *optrA* plasmid (pE394, accession KP399637), with only the 5–10 kb region surrounding *optrA* and *fexA* (a chloramphenicol/florfenicol exporter) showing >70% nucleotide identity. In all six Scottish *optrA*-positive plasmids, *optrA* and *fexA* were located within 550–750 nt of each other intervened by a single coding sequence (hypothetical function in all but pBX8117-2 which was annotated as a putative NADH reductase). Within the Scottish *optrA*-positive plasmids, pBX5936-1 (69 kb) and pTM6294-2 (53 kb) were most similar, sharing 97% average nucleotide identity over 40 kb of aligned sequence (Fig. 1). pTM6294-2 shared 99.8% identity with a 53 kb *optrA*-positive pheromone responsive plasmid detected in *E. faecalis* from a clinical sample in China (pEF10748), clinical samples in Spain (IsoBar1, IsoBar2 and IsoBar3) and raw dog food in Portugal (pAPT110) [37, 38]. pWE0438 shared 92.3% nucleotide identity over 52 kb with pS7316 from an *E. faecalis* isolated from a hospitalized patient in Japan [39]. In pWE0438, the *optrA* and *fexA* genes were ~3.8 kb upstream of Tn*917* carrying *ermB*, and ~1.8 kb downstream of another Tn*3*-family transposase (Fig. 1). pBX8117-2 carried *optrA* and the novel *cfr*(D) gene (encoding a 23S rRNA methylase that confers phenicol, oxazolidinone, pleuromutilin and streprogramin A resistance) but apart from these genes showed no similarity to another *E. faecium optrA*/*cfr*(D)-positive plasmid identified in a clinical sample in Ireland (M17-0314) [40]. The other Scottish *optrA*-positive plasmids showed limited similarity to other published examples outside of the *optrA*/*fexA* region.

IS*1216* is often associated with *optrA* and other antimicrobial resistance genes in enterococci. pBX5936-1 and pBX8117-2 had IS*1216* flanking the *optrA* and *fexA* region as a putative transposable cassette (Figs 1 and S1). However, IS*1216* can mobilize from a single insertion sequence copy [41] and single copies were found close to *optrA* in pTM6294-2 and pWE0851-1 (Figs 1 and S1). BLASTN comparison of pWE0254-1 with the other *optrA*-positive plasmids highlighted a partial IS*1216* transposase that was not identified by automated annotation. Immediately upstream of the partial IS*1216* was an IS*3*-family transposase, the insertion of which probably disrupted IS*1216*. In pWE0254-1 *optrA* and *fexA* were found on a Tn*6674*-like element carrying macrolide (*ermA*) and spectinomycin (APH(9)-Ia) resistance genes. The element was 98.9% identical to Tn*6674* but had a 1.2 kb insertion containing IS*3*-family transposases (Fig. S1), and was classified as Tn*6993* by the Transposon Registry (accession GCA_906464915) [42]. Tn*6993* was not inserted into the chromosomal *radC* gene as described for most Tn*6674*-like elements [43, 44]. A similar element was present in a plasmid from *E. faecalis* in Chinese swine (TBCP-4814-p1, accession MH830363) but this element lacked the *tnpA* gene and the 1.2 kb insertion of Tn*6993* (Fig. S1) [45]. pWE0438 had a single copy of IS*1216* located ~35 kb from *optrA*, although Tn*917* and Tn*3*-like transposases were detected closer to *optrA* as described above.

optrA sequences vary between isolates

Comparison of the *optrA* sequence from each isolate to the first identified *optrA* from pE394 revealed different variants at the nucleotide and amino acid levels: WE0254 and TM6294 had one synonymous nucleotide substitution, BX5936 had a single non-synonymous nucleotide substitution, WE0851 had two non-synonymous nucleotide substitutions, WE0348 had three non-synonymous and one synonymous substitution, and BX8117 had 20 non-synonymous and a further 17 synonymous substitutions (Table S1, Fig. S2). The degree of sequence variation between the six FexA proteins was less than that seen in OptrA. Comparison

Table 1. Details of the *optrA*-positive *E. faecalis* characterized in this study

Isolate	Year	Region	Clinical sample	Patient source	MLST	Acquired linezolid resistance genes					Mutations in 23S rRNA		Mutations in ribosomal proteins*			MIC (mg l^{-1})	
						cfr	*cfr*(B)	*cfr*(D)	o*ptrA*	*poxtA*	G2505A	G2576T	L3	L4	L22	CHL	LZD
WE0851	2014	Grampian	Urine	GP	480	–	–	–	+	–	–	–	T150A	F101L	–	≥64	8
WE0254	2015	Grampian	Urine	GP	19	–	–	–	+	–	–	–	T150A	F101L	–	≥64	8
WE0438	2016	Grampian	Urine	Hospital	330	–	–	–	+	–	–	–	T150A	F101L	–	≥64	8
TM6294	2017	Forth Valley	Urine	Hospital	585	–	–	–	+	–	–	–	T150A	F101L	–	≥64	8
BX5936	2017	Lothian	Semen	GP	894	–	–	–	+	–	–	–	T150A	F101L	–	≥64	8
BX8117	2017	Lothian	Urine	GP	16	–	–	+	+	–	–	–	T150A	F101L	–	≥64	8

GP, general practice.
*The mutations identified here have never been detected in the absence of other resistance mechanisms in linezolid-resistant isolates, and have been detected in linezolid-susceptible isolates. Their role in linezolid resistance is unclear [54].

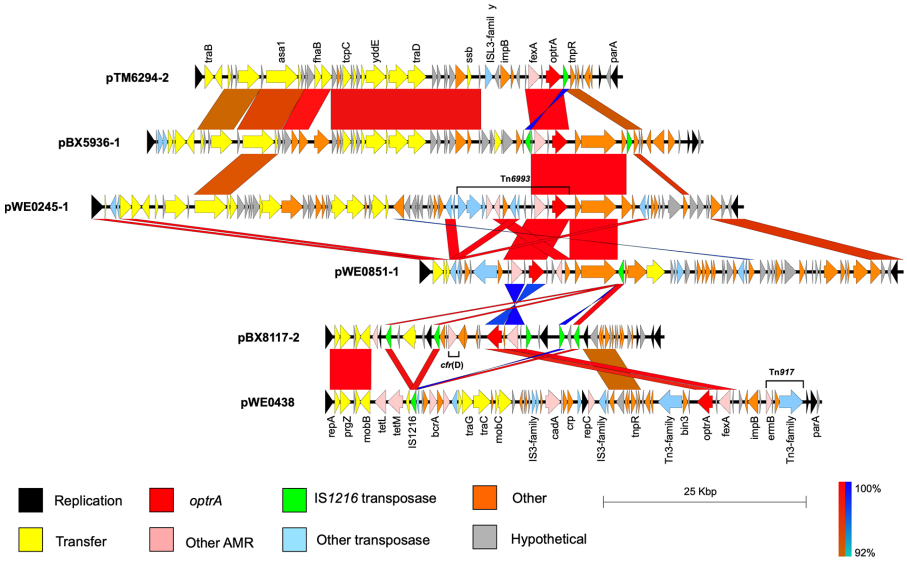


Fig. 1. Alignment of full *optrA*-positive plasmid sequences. While some sequence similarity is seen between pTM6294-2 and pBX5936-1, in general identity is low between the *optrA*-positive plasmids, indicating *optrA* has mobilized to multiple plasmid backbones. Arrows indicate coding sequences, coloured blocks between each sequence indicate regions with BLASTN sequence identity ≥90% and length >680 bp. Blue identity blocks indicate inverted sequence.

with the first reported FexA sequence (AJ549214) showed four common non-synonymous variants in all strains (amino acid changes A34S, L39S, I131V and V305I), with all but BX8117 having an additional D50A variant.

DISCUSSION

This study found *optrA* present in diverse genetic lineages of *E. faecalis* and carried on largely unrelated plasmids in six isolates from Scotland. pTM6294-2, pBX5936-1 and pWE0438 shared homology with plasmids identified in China or Japan, highlighting the wide dispersal of *optrA*. However, the other Scottish plasmids had limited similarity to other published examples, suggesting a diverse reservoir of *optrA*-carrying genetic elements. We identified *optrA* often carried with a number of other resistance genes, including in a novel multiresistance transposon Tn*6993* in pWE0254-1, and the recently described *cfr*(D) in pBX8117-2. Despite differences in *optrA* sequences and carriage of other linezolid determinants such as *cfr*(D), all isolates showed low-level linezolid resistance of 8 mg l^{-1} (Table 1).

Freitas *et al.* [44] recently analysed all publicly available *optrA*-positive genome sequences and categorized the genetic environment of *optrA*. Group 1 includes Tn*6674*-like platforms, of which WE0254 is a representative (Fig. S1). However, in the original scheme all Group 1 elements were integrated into the chromosome, while in WE0254 the *optrA* element Tn*6993* is inserted into a plasmid. Group 2 includes *optrA-fexA-impB* platforms, represented in the Scottish isolates by TM6294 and WE0438 (Fig. S1). Group 3 includes platforms containing the *araC* regulatory element and is not represented in the Scottish *optrA*-positive isolates characterized here. The three remaining Scottish isolates could not be grouped based on the Freitas scheme, highlighting the need for further studies and public access to complete genome sequences to determine the true diversity of *optrA*-positive platforms.

Many studies of *optrA* to date are from China and tend to show a higher prevalence of *optrA* in isolates from animals rather than humans [11, 46, 47]. Additionally, florfenicol use in agriculture is linked to *optrA* detection in farm animals [48, 49]. However, increasing reports describe rapid increases in *optrA* detection from human samples in many countries [15, 50, 51]. *optrA*-positive isolates are often resistant to multiple antibiotic classes used in animal and human health, allowing significant opportunity for co-selection of *optrA*-positive strains both in animal and in human settings. More recently, *optrA* has been identified in clinical vancomycin-resistant *E. faecium* isolates, with very limited treatment options [50, 52, 53].

Our study is limited in scale as we only include isolates from three regional clinical laboratories, and therefore larger studies are required to infer national patterns. However, our finding that *optrA* is present as different gene variants, carried on different mobile genetic elements, in unrelated strains of *E. faecalis* suggest a diverse *optrA* reservoir that is only partly investigated in this study.

As well as *optrA*, the *cfr* and *poxtA* genes are emerging transferable linezolid resistance mechanisms. Further studies from a One Health perspective are warranted to understand the selection pressures driving transferable linezolid resistance, and the transmission dynamics of these strains to avoid further spread of oxazolidinone resistance within *E. faecalis* and other Gram-positive bacteria.

Funding information
This work was supported by the Chief Scientist Office (Scotland) through the Scottish Healthcare Associated Infection Prevention Institute (Reference SIRN/10). Bioinformatics and Computational Biology analyses were supported by the University of St Andrews Bioinformatics Unit, which is funded by a Wellcome Trust ISSF award [grant 105621/Z/14/Z].

Acknowledgements
The authors would like to thank the Bioinformatics Unit at the University of St Andrews and Pathogen Informatics at the Wellcome Sanger Institute for access to high-performance computing clusters.

Conflicts of interest
The authors declare that there are no conflicts of interest.

Ethical statement
Access to isolates and clinical data was approved by the NHS Scotland Biorepository Network (Ref TR000126).

References

1. García-Solache M, Rice LB. The Enterococcus: a model of adaptability to its environment. *Clin Microbiol Rev* 2019;32:e00058-18.
2. ECDC. Surveillance of antimicrobial resistance in Europe – Annual report of the European Antimicrobial Resistance Surveillance Network (EARS-Net) 2017; 2019. http://www.ecdc.europa.eu/sites/portal/files/documents/EARS-Net-report-2017-update-jan-2019.pdf
3. Zahedi Bialvaei A, Rahbar M, Yousefi M, Asgharzadeh M, Samadi Kafil H. Linezolid: a promising option in the treatment of Gram-positives. *J Antimicrob Chemother* 2017;72:354–364.
4. Public Health England. English surveillance programme for antimicrobial utilisation and resistance (ESPAUR) Report 2018-2019. London, UK: PHE; 2019. https://www.gov.uk/government/publications/english-surveillance-programme-antimicrobial-utilisation-and-resistance-espaur-report
5. Health Protection Scotland. Scottish One Health Antimicrobial Use and Resistance in 2018 Annual Report. Glasgow, UK: HPS; 2019. https://www.hps.scot.nhs.uk/web-resources-container/scottish-one-health-antimicrobial-use-and-antimicrobial-resistance-in-2018
6. Mendes RE, Deshpande LM, Jones RN. Linezolid update: stable in vitro activity following more than a decade of clinical use and summary of associated resistance mechanisms. *Drug Resist Updat* 2014;17:1–12.
7. Abbo L, Shukla BS, Giles A, Aragon L, Jimenez A, *et al*. Linezolid- and Vancomycin-resistant *Enterococcus faecium* in solid organ transplant recipients: infection control and antimicrobial stewardship using whole genome sequencing. *Clin Infect Dis* 2019;69:259–265.
8. Antonelli A, D'Andrea MM, Brenciani A, Galeotti CL, Morroni G, *et al*. Characterization of *poxtA*, a novel phenicol-oxazolidinone-tetracycline resistance gene from an MRSA of clinical origin. *J Antimicrob Chemother* 2018;73:1763–1769.
9. Deshpande LM, Ashcraft DS, Kahn HP, Pankey G, Jones RN, *et al*. Detection of a New *cfr*-like gene, cfr(B), in *Enterococcus faecium* isolates recovered from human specimens in the United States as Part of the SENTRY Antimicrobial Surveillance Program. *Antimicrob Agents Chemother* 2015;59:6256–6261.
10. Diaz L, Kiratisin P, Mendes RE, Panesso D, Singh KV, *et al*. Transferable plasmid-mediated resistance to linezolid due to cfr in a human clinical isolate of *Enterococcus faecalis*. *Antimicrob Agents Chemother* 2012;56:3917–3922.
11. Wang Y, Lv Y, Cai J, Schwarz S, Cui L, *et al*. A novel gene, *optrA*, that confers transferable resistance to oxazolidinones and phenicols and its presence in *Enterococcus faecalis* and *Enterococcus faecium* of human and animal origin. *J Antimicrob Chemother* 2015;70:2182–2190.
12. Pang S, Boan P, Lee T, Gangatharan S, Tan SJ, *et al*. Linezolid-resistant ST872 *Enteroccocus faecium* harbouring *optrA* and *cfr* (D) oxazolidinone resistance genes. *Int J Antimicrob Agents* 2020;55:105831.
13. Long KS, Poehlsgaard J, Kehrenberg C, Schwarz S, Vester B. The Cfr rRNA methyltransferase confers resistance to Phenicols, Lincosamides, Oxazolidinones, Pleuromutilins, and Streptogramin A antibiotics. *Antimicrob Agents Chemother* 2006;50:2500–2505.
14. Guerin F, Sassi M, Dejoies L, Zouari A, Schutz S, *et al*. Molecular and functional analysis of the novel cfr(D) linezolid resistance gene identified in *Enterococcus faecium*. *J Antimicrob Chemother* 2020;75:1699–1703.
15. Deshpande LM, Castanheira M, Flamm RK, Mendes RE. Evolving oxazolidinone resistance mechanisms in a worldwide collection of enterococcal clinical isolates: results from the SENTRY Antimicrobial Surveillance Program. *J Antimicrob Chemother* 2018;73:2314–2322.
16. Health Protection Scotland. Oxazolidinone-resistance due to *optrA* in *Enterococcus faecalis*. *HPS Wkly Rep* 2016;50:230–231.
17. Cai J, Wang Y, Schwarz S, Zhang G, Chen S, *et al*. High detection rate of the oxazolidinone resistance gene *optrA* in *Enterococcus faecalis* isolated from a Chinese anorectal surgery ward. *Int J Antimicrob Agents* 2016;48:757–759.
18. He T, Shen Y, Schwarz S, Cai J, Lv Y, *et al*. Genetic environment of the transferable oxazolidinone/phenicol resistance gene *optrA* in *Enterococcus faecalis* isolates of human and animal origin. *J Antimicrob Chemother* 2016;71:1466–1473.
19. EUCAST. *Breakpoint tables for interpretation of MICs and zone diameters*. Version 8.0. *Version* 2018.
20. Woodford N, Tysall L, Auckland C, Stockdale MW, Lawson AJ, *et al*. Detection of oxazolidinone-resistant *Enterococcus faecalis* and *Enterococcus faecium* strains by real-time PCR and PCR-restriction fragment length polymorphism analysis. *J Clin Microbiol* 2002;40:4298–4300.
21. Werner G, Strommenger B, Klare I, Witte W. Molecular detection of linezolid resistance in *Enterococcus faecium* and *Enterococcus faecalis* by use of 5' nuclease real-time PCR compared to a modified classical approach. *J Clin Microbiol* 2004;42:5327–5331.

22. **Kehrenberg C**, **Schwarz S**. Distribution of florfenicol resistance genes *fexA* and *cfr* among chloramphenicol-resistant *Staphylococcus* isolates. *Antimicrob Agents Chemother* 2006;50:1156–1163.
23. **Bolger AM**, **Lohse M**, **Usadel B**. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30:2114–2120.
24. **Loman NJ**, **Quinlan AR**. Poretools: a toolkit for analyzing nanopore sequence data. *Bioinformatics* 2014;30:3399–3401.
25. **Ponstingl H**, **Ning Z**. SMALT. *Wellcome Trust Sanger Institute*; 2014. http://www.sanger.ac.uk/science/tools/smalt-0 [accessed 21 June 2017].
26. **Page AJ**, **Taylor B**, **Delaney AJ**, **Soares J**, **Seemann T**, ***et al***. *SNP-sites*: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb Genom* 2016;2:e000056.
27. **Inouye M**, **Dashnow H**, **Raven L-A**, **Schultz MB**, **Pope BJ**, ***et al***. SRST2: rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med* 2014;6:1–16.
28. **Jolley KA**, **Maiden MCJ**. BIGSdb: scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* 2010;11:595.
29. **Ruiz-Garbajosa P**, **Bonten MJM**, **Robinson DA**, **Top J**, **Nallapareddy SR**, ***et al***. Multilocus sequence typing scheme for *Enterococcus faecalis* reveals hospital-adapted genetic complexes in a background of high rates of recombination. *J Clin Microbiol* 2006;44:2220–2228.
30. **Hunt M**, **Mather AE**, **Sánchez-Busó L**, **Page AJ**, **Parkhill J**, ***et al***. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb Genom* 2017;3:e000131.
31. **Zankari E**, **Hasman H**, **Kaas RS**, **Seyfarth AM**, **Agersø Y**, ***et al***. Genotyping using whole-genome sequencing is a realistic alternative to surveillance based on phenotypic antimicrobial susceptibility testing. *J Antimicrob Chemother* 2013;68:771–777.
32. **Wick RR**, **Judd LM**, **Gorrie CL**, **Holt KE**. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 2017;13:e1005595.
33. **Pruitt KD**, **Tatusova T**, **Brown GR**, **Maglott DR**. NCBI reference sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res* 2012;40:D130-5.
34. **Sullivan MJ**, **Petty NK**, **Beatson SA**. Easyfig: a genome comparison visualizer. *Bioinformatics* 2011;27:1009–1010.
35. **Raven KE**, **Reuter S**, **Gouliouris T**, **Reynolds R**, **Russell JE**, ***et al***. Genome-based characterization of hospital-adapted *Enterococcus faecalis* lineages. *Nat Microbiol* 2016;1:15033.
36. **Goodhead I**, **Darby AC**. Taking the pseudo out of pseudogenes. *Curr Opin Microbiol* 2015;23:102–109.
37. **Zou J**, **Tang Z**, **Yan J**, **Liu H**, **Chen Y**, ***et al***. (n.d.) Dissemination of linezolid resistance through sex pheromone plasmid transfer in *Enterococcus faecalis*. *Front Microbiol*;11.
38. **Freitas AR**, **Finisterra L**, **Tedim AP**, **Duarte B**, **Novais C**, ***et al***. Linezolid- and multidrug-resistant enterococci in raw commercial dog foodMultidrug-Resistant Enterococci in Raw Commercial Dog Food, Europe, 2019-2020. *Emerg Infect Dis* 2021;27:2221–2224.
39. **Iimura M**, **Hayashi W**, **Arai E**, **Natori T**, **Horiuchi K**, ***et al***. Identification of a multiresistant mosaic plasmid carrying a new segment of IS1216E-flanked *optrA* with integrated Tn551-ermB element in linezolid-resistant *Enterococcus faecalis* human isolate. *J Glob Antimicrob Resist* 2020;22:697–699.
40. **Egan SA**, **Shore AC**, **O'Connell B**, **Brennan GI**, **Coleman DC**. Linezolid resistance in *Enterococcus faecium* and *Enterococcus faecalis* from hospitalized patients in Ireland: high prevalence of the MDR genes *optrA* and poxtA in isolates with diverse genetic backgrounds. *J Antimicrob Chemother* 2020;75:1704–1711.
41. **Partridge SR**, **Kwong SM**, **Firth N**, **Jensen SO**. Mobile genetic elements associated with antimicrobial resistance. *Clin Microbiol Rev* 2018;31:e00088-17.
42. **Tansirichaiya S**, **Rahman MA**, **Roberts AP**. The transposon registry. *Mob DNA* 2019;10:40.
43. **Li D**, **Li X-Y**, **Schwarz S**, **Yang M**, **Zhang S-M**, ***et al***. Tn *6674*, a novel enterococcal *optrA* -carrying multiresistance transposon of the TN *554* family. *Antimicrob Agents Chemother* 2019:AAC.00809-19, aac;AAC.00809-19v1.
44. **Freitas AR**, **Tedim AP**, **Novais C**, **Lanza VF**, **Peixe L**. Comparative genomics of global *optrA*-carrying *Enterococcus faecalis* uncovers a common chromosomal hotspot for *optrA* acquisition within a diversity of core and accessory genomes. *Microb Genom* 2020;6.
45. **Chen L**, **Han D**, **Tang Z**, **Hao J**, **Xiong W**, ***et al***. Co-existence of the oxazolidinone resistance genes cfr and *optrA* on two transferable multi-resistance plasmids in one *Enterococcus faecalis* isolate from swine. *Int J Antimicrob Agents* 2020;56:105993.
46. **Shang Y**, **Li D**, **Hao W**, **Schwarz S**, **Shan X**, ***et al***. A prophage and two ICESa2603-family integrative and conjugative elements (ICEs) carrying *optrA* in *Streptococcus suis*. *J Antimicrob Chemother* 2019;74:2876–2879.
47. **Hao W**, **Shan X**, **Li D**, **Schwarz S**, **Zhang S-M**, ***et al***. Analysis of a poxtA- and *optrA*-co-carrying conjugative multiresistance plasmid from Enterococcus faecalis. *J Antimicrob Chemother* 2019;74:1771–1775.
48. **Munk P**, **Knudsen BE**, **Lukjancenko O**, **Duarte ASR**, ***et al***. Abundance and diversity of the faecal resistome in slaughter pigs and broilers in nine European countries. *Nat Microbiol* 2018;3:898–908.
49. **Zhao Q**, **Wang Y**, **Wang S**, **Wang Z**, **Du X**, ***et al***. Prevalence and abundance of florfenicol and linezolid resistance genes in soils adjacent to swine feedlots. *Sci Rep* 2016;6:1–7.
50. **Sassi M**, **Guérin F**, **Zouari A**, **Beyrouthy R**, **Auzou M**, ***et al***. Emergence of *optrA*-mediated linezolid resistance in enterococci from France, 2006-16. *J Antimicrob Chemother* 2019;74:1469–1472.
51. **Bender JK**, **Fleige C**, **Lange D**, **Klare I**, **Werner G**. Rapid emergence of highly variable and transferable oxazolidinone and phenicol resistance gene *optrA* in German *Enterococcus* spp. clinical isolates. *Int J Antimicrob Agents* 2018;52:819–827.
52. **Egan SA**, **Corcoran S**, **McDermott H**, **Fitzpatrick M**, **Hoyne A**, ***et al***. Hospital outbreak of linezolid-resistant and vancomycin-resistant ST80 Enterococcus faecium harbouring an *optrA*-encoding conjugative plasmid investigated by whole-genome sequencing. *J Hosp Infect* 2020;105:726–735.
53. **Lazaris A**, **Coleman DC**, **Kearns AM**, **Pichon B**, **Kinnevey PM**, ***et al***. Novel multiresistance cfr plasmids in linezolid-resistant methicillin-resistant *Staphylococcus epidermidis* and vancomycin-resistant *Enterococcus faecium* (VRE) from a hospital outbreak: co-location of cfr and *optrA* in VRE. *J Antimicrob Chemother* 2017;72:3252–3257.
54. **Cui L**, **Wang Y**, **Lv Y**, **Wang S**, **Song Y**, ***et al***. Nationwide surveillance of novel oxazolidinone resistance gene *optrA* in *Enterococcus Isolates* in China from 2004 to 2014. *Antimicrob Agents Chemother* 2016;60:7490–7493.