Antibiotic resistance determination using *Enterococcus faecium* whole-genome sequences: a diagnostic accuracy study using genotypic and phenotypic data

Francesc Coll, Theodore Gouliouris, Beth Blane, Corin A Yeats, Kathy E Raven, Catherine Ludden, Fahad A Khokhar, Hayley J Wilson, Leah W Roberts, Ewan M Harrison, Carolyne S Horner, Le Thi Hoi, Nguyen Thi Hoa, Nguyen Vu Trung, Nicholas M Brown, Mark A Holmes, Julian Parkhill, Mili Estee Török, Sharon J Peacock

Summary

Background DNA sequencing could become an alternative to in vitro antibiotic susceptibility testing (AST) methods for determining antibiotic resistance by detecting genetic determinants associated with decreased antibiotic susceptibility. Here, we aimed to assess and improve the accuracy of antibiotic resistance determination from *Enterococcus faecium* genomes for diagnosis and surveillance purposes.

Methods In this retrospective diagnostic accuracy study, we first conducted a literature search in PubMed on Jan 14, 2021, to compile a catalogue of genes and mutations predictive of antibiotic resistance in *E faecium*. We then evaluated the diagnostic accuracy of this database to determine susceptibility to 12 different, clinically relevant antibiotics using a diverse population of 4382 *E faecium* isolates with available whole-genome sequences and in vitro culture-based AST phenotypes. Isolates were obtained from various sources in 11 countries worldwide between 2000 and 2018. We included isolates tested with broth microdilution, Vitek 2, and disc diffusion, and antibiotics with at least 50 susceptible and 50 resistant isolates. Phenotypic resistance was derived from raw minimum inhibitory concentrations and measured inhibition diameters, and harmonised primarily using the breakpoints set by the European Committee on Antimicrobial Susceptibility Testing. A bioinformatics pipeline was developed to process raw sequencing reads, identify antibiotic resistance genetic determinants, and report genotypic resistance. We used our curated database, as well as ResFinder, AMRFinderPlus, and LRE-Finder, to assess the accuracy of genotypic predictions against phenotypic resistance.

Findings We curated a catalogue of 228 genetic markers involved in resistance to 12 antibiotics in *E faecium*. Very accurate genotypic predictions were obtained for ampicillin (sensitivity 99·7% [95% CI 99·5–99·9] and specificity 97·9% [95·8–99·0]), ciprofloxacin (98·0% [96·4–98·9] and 98·8% [95·9–99·7]), vancomycin (98·8% [98·3–99·2] and 98·8% [98·0–99·3]), and linezolid resistance (after re-testing false negatives: 100·0% [90·8–100·0] and 98·3% [97·8–98·7]). High sensitivity was obtained for tetracycline (99·5% [99·1–99·7]), teicoplanin (98·9% [98·4–99·3]), and high-level resistance to aminoglycosides (97·7% [96·6–98·4] for streptomycin and 96·8% [95·8–97·5] for gentamicin), although at lower specificity (60–90%). Sensitivity was expectedly low for daptomycin (73·6% [65·1–80·6]) and tigecycline (38·3% [27·1–51·0]), for which the genetic basis of resistance is not fully characterised. Compared with other antibiotic resistance databases and bioinformatic tools, our curated database was similarly accurate at detecting resistance to ciprofloxacin and linezolid and high-level resistance to streptomycin and gentamicin, but had better sensitivity for detecting resistance to ampicillin, tigecycline, daptomycin, and quinupristin–dalfopristin, and better specificity for ampicillin, vancomycin, teicoplanin, and tetracycline resistance. In a validation dataset of 382 isolates, similar or improved diagnostic accuracies were also achieved.

Interpretation To our knowledge, this work represents the largest published evaluation to date of the accuracy of antibiotic susceptibility predictions from *E faecium* genomes. The results and resources will facilitate the adoption of whole-genome sequencing as a tool for the diagnosis and surveillance of antimicrobial resistance in *E faecium*. A complete characterisation of the genetic basis of resistance to last-line antibiotics, and the mechanisms mediating antibiotic resistance silencing, are needed to close the remaining sensitivity and specificity gaps in genotypic predictions.

Funding Wellcome Trust, UK Department of Health, British Society for Antimicrobial Chemotherapy, Academy of Medical Sciences and the Health Foundation, Medical Research Council Newton Fund, Vietnamese Ministry of Science and Technology, and European Society of Clinical Microbiology and Infectious Disease.

Lancet Microbe 2024; 5: e151–63

Published Online January 11, 2024 https://doi.org/10.1016/S2666-5247(23)00297-5

Department of Infection Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London, UK (F Coll PhD); **Parasites & Microbes Programme, Wellcome Sanger Institute, Hinxton, Cambridge, UK** (F Coll, E M Harrison PhD); **Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK** (T Gouliouris MD PhD, N M Brown PhD FRCP, M Estee Török PhD FRCP, Prof S J Peacock PhD FRCP); **Department of Medicine** (T Gouliouris, B Blane MSc, K E Raven PhD, L W Roberts PhD, E M Harrison, Prof S J Peacock) **and Department of Veterinary Medicine** (F A Khokhar BSc, H J Wilson PhD, Prof M A Holmes PhD, Prof J Parkhill PhD), **University of Cambridge, Cambridge, UK; Centre for Genomic Pathogen Surveillance, Big Data Institute, Nuffield Department of Medicine, University of Oxford, Oxford, UK** (C A Yeats PhD); **Genomics England, London, UK** (C Ludden PhD); **European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Hinxton, UK** (L W Roberts); **British Society for Antimicrobial Chemotherapy, Birmingham, UK** (C S Horner PhD, N M Brown); **National Hospital for Tropical Diseases, Hanoi, Viet Nam** (T H Le MD, T H Nguyen MD, V T Nguyen MD); **Hanoi Medical University, Hanoi, Viet Nam** (T H Le, V T Nguyen); **Department of Microbiology and National Tuberculosis Reference Laboratory, National Lung**

Hospital, Hanoi, Viet Nam (T H Nguyen)**; UK Health Security Agency, Cambridge, UK** (N M Brown)

Correspondence to:
Dr Francesc Coll, Department of Infection Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK
francesc.coll@lshtm.ac.uk

or

Dr Theodore Gouliouris, Cambridge University Hospitals NHS Foundation Trust, Cambridge CB2 0QQ, UK
theo.gouliouris@doctors.org.uk

Introduction

WHO has declared antimicrobial resistance (AMR) as one of the top ten global public health threats facing humanity. Without effective antimicrobials, the success of modern medicine in treating infections, including those acquired during major surgery and cancer chemotherapy, would be at increased risk. Antibiotic-resistant strains of *Enterococcus faecium*, especially vancomycin-resistant *E faecium*, are of particular concern. *E faecium* is a leading cause of hospital-acquired bloodstream, surgical site, or urinary tract infections, especially in immunocompromised and critically ill patients. Health-care-associated strains of *E faecium* are commonly resistant to multiple antibiotics, including ampicillin and vancomycin. Antibiotic-resistant *E faecium* is among the top ten bacterial pathogens with the highest AMR-associated mortality burden, with an estimated 200 000 deaths per year globally.[1] As such, vancomycin-resistant *E faecium* is regarded as a high priority pathogen by WHO.[2] Antibiotic susceptibility testing (AST) is needed to ensure that the right antibiotics are prescribed to treat bacterial infections, and is routinely performed using culture-based methods such as disc diffusion or broth microdilution.

As antibiotic resistance is genetically encoded, whole-genome sequencing has emerged as an attractive tool for the detection of AMR as it can, in principle, be used to predict resistance to virtually all antibiotics in a single experiment.[3] The accuracy of such predictions depends on the availability of curated and updated databases of genetic determinants of AMR, and large collections of isolates with AST phenotypes and genome sequences to assess the diagnostic performance of such databases. Automated bioinformatics tools are also needed to analyse and interpret genomic data.[4,5]

The success in predicting phenotypic resistance from whole-genome sequences is well exemplified for other bacterial species such as *Mycobacterium tuberculosis*,[6] *Staphylococcus aureus*,[7] or *Streptococcus pneumoniae*.[8] For *E faecium*, previous studies[3,9–13] used isolate collections of small sample sizes (100–205 isolates) to assess genotype–phenotype concordance for seven to 13 antibiotics, and with relatively low resistance rates for some antibiotics.

In this study, we aimed to assess and improve the accuracy of AMR determination from *E faecium* genomes by assembling a catalogue of acquired genes and mutations predictive of AMR in *E faecium*, and evaluating the diagnostic accuracy of genotypic predictions obtained with such a catalogue against culture-based antibiotic susceptibility phenotypes.

Methods

Study design

In this retrospective diagnostic accuracy study, we assembled a collection of 4382 *E faecium* isolates with whole-genome sequences and AMR phenotypes available with

Research in context

Evidence before this study

Early proof-of-concept studies showed the feasibility of determining antimicrobial resistance (AMR) from bacterial genome sequences. Population-based studies are needed to assess the diagnostic accuracy of genotypic predictions. We conducted a literature search to identify studies using whole-genome sequencing (WGS) to detect AMR in *Enterococcus faecium*. We searched PubMed using the terms "*Enterococcus faecium*"; "whole genome sequencing" or "WGS"; "antimicrobial"; "resistance" or "susceptibility"; and "prediction", "determination", or "diagnosis" for articles published up to Dec 22, 2022. Of the 161 articles retrieved, five investigated the concordance between WGS-based predictions and phenotypic antibiotic susceptibility testing (AST) results. One study investigated the prediction of linezolid resistance specifically. The rest considered seven to 13 antibiotics, using relatively modest sample sizes (100–205 isolates), and low numbers (<15) of resistant (eg, to tigecycline) or susceptible isolates (eg, to ampicillin) for some antibiotics. ResFinder, LRE-finder, or custom catalogues of resistance markers were used, followed by expert interpretation of the results, to establish concordance between genotypic and phenotypic AST results.

Added value of this study

We curated a catalogue of 228 genetic markers involved in resistance to 12 clinically relevant different antibiotics in *E faecium*. We used a collection of 4382 *E faecium* isolates with whole-genome sequences and AST phenotypes available to assess the diagnostic accuracy of this catalogue. Nearly complete genotype–phenotype concordance could be achieved for ampicillin, ciprofloxacin, vancomycin, and linezolid. High sensitivity was also obtained for detecting resistance to tetracyclines and teicoplanin and high-level resistance to aminoglycosides, although with lower specificity than that achieved for other antibiotics. Sensitivity was lower for the detection of daptomycin and tigecycline resistance than for the other antibiotics tested. We additionally assessed ResFinder and AMRFinderPlus and found an overall improved accuracy of our curated database compared with these tools and databases. We have made our curated database of AMR genetic markers, isolates' whole-genome sequences and AST data, and underlying bioinformatic pipelines publicly available.

Implications of all the available evidence

Our study represents an important improvement in the accuracy of AMR determination from *E faecium* genomes with respect to existing AMR detection tools and databases, and provides specific future directions to close the remaining sensitivity and specificity gaps in genotypic predictions. Given the mortality burden of AMR *E faecium* infection and the increasingly routine use of WGS in clinical microbiology laboratories, the results and resources made available here will facilitate the adoption of WGS as a tool for the diagnosis and surveillance of antimicrobial resistance in *E faecium*.

the goal of determining the concordance between genotypic and phenotypic AMR in *E faecium*. The isolates we used had been obtained from 11 countries worldwide (Australia, Austria, Canada, Denmark, Germany, the Netherlands, New Zealand, Pakistan, the UK, the USA, and Viet Nam) between 2000 and 2018. We extracted previously published and newly available minimum inhibitory concentrations (MICs) and disc diameter measurements used to define resistance for selected antibiotics and applied the same breakpoints to harmonise phenotypic resistance across isolates. Bacterial genomes were interrogated for the presence of genes and mutations predictive of AMR in *E faecium*, as curated from a literature review or identified by other available AMR databases. Using phenotypic resistance as the reference standard, the accuracy of genotypic predictions was calculated. No ethics approval was required for this study as it involved secondary analysis of bacterial genomic data.

Collections of whole-genome-sequenced strains

We used publicly available collections of *E faecium* from published studies[3,9,11,12,14–26] for which whole-genome sequences and antibiotic susceptibility phenotypes were available. Only studies with AST data from broth microdilution, Vitek 2 (bioMérieux; an automated bacterial identification and AST system), or disc diffusion were used in this study. Data obtained from E-test (bioMérieux) and similar gradient test strips were excluded, resulting in 947 isolates obtained from publicly available collections.[3,9,11,12,15,16,18–21,23,24] We also used the phenotypic AST data of an additional 3435 *E faecium* isolates whose genome sequences had already been published by coauthors,[27–32] who made the AST data available for this study, and from an unpublished study from coauthors conducted in the intensive-care unit of Addenbrookes Hospital, Cambridge, UK. We included 12 clinically relevant antibiotics that had at least 50 susceptible and 50 resistant isolates available across studies: ampicillin, ciprofloxacin, daptomycin, doxycycline, gentamicin, streptomycin, linezolid, quinupristin–dalfopristin, tetracycline, tigecycline, teicoplanin, and vancomycin. We also considered clindamycin, erythromycin, and kanamycin, which are not commonly used to treat enterococci infections. Additionally, we used a published collection of 382 *E faecium* isolates as an independent and validation dataset (appendix 1, tab 1).[10] See appendix 2 (pp 3–4) for details on extraction and standardisation of phenotypic AST data. Our preference was to use clinical breakpoints from the European Committee on Antimicrobial Susceptibility Testing (EUCAST) version 12.0 (appendix 2 pp 17–18) to define resistance as they resulted in better phenotype–genotype concordance than breakpoints from the Clinical and Laboratory Standards Institute (CLSI; appendix 2 pp 9–10, 19).

Literature review on genetic determinants of AMR

We started by extracting a list of antibiotics and AMR-associated genes from reviews and AMR genotype–phenotype studies in enterococci (appendix 2 p 4). We then conducted a targeted PubMed search on Jan 14, 2021, for articles published in English, using the search terms ("enterococcus faecium" OR "E. faecium" OR "enterococci") AND (resistan* OR susceptib* OR sensitiv*) AND (antibiotic) AND (gene name), for which the terms "antibiotic" and "gene name" were specified for each antibiotic–gene combination to identify the earliest and cumulative evidence on the role of each chromosomal and acquired gene in the susceptibility to individual antibiotics, not to antibiotic classes. For cases of mutational resistance, we added the PubMed search terms "AND (mutation* OR polymorphism* OR SNP*)" to capture all reported mutations, and combinations of mutations, in chromosomal AMR-associated genes. Promoter regions of AMR-associated genes were not analysed in this study. See appendix 2 (pp 4–5) for details.

In-silico prediction of AMR from whole-genome sequences

We used ARIBA (Antimicrobial Resistance Identification By Assembly) software for the identification of antibiotic resistance genes and mutations from whole-genome sequences. We formatted our curated database to ARIBA-compliant input files using a custom python script (prepare_ariba_files_from_amr_database.py). ARIBA version 2.14.6 was run separately on each isolate genome. For isolates with only genome assemblies available, perfect paired-end fastq reads were generated from assemblies using fastaq version 3.17.0. A custom python script was written to parse ARIBA reports and call resistance (call_resistance_from_ariba_report.py) on the basis of the presence of single or combinations of mutations, and on the presence and integrity (ie, fully assembled genes without frameshifts or truncations) of single or multiple acquired genes. This script uses a default threshold of 90% nucleotide identity and a 60% threshold of minimum gene length to detect assembled AMR-associated genes from ARIBA reports. It also includes the option (recommended) to use ARIBA flags to establish whether genes are labelled as assembled. A minimum ratio of reads supporting alternative alleles of 10% to predict resistance from heterozygous mutations was used, as previously proposed.[9] This script also includes the option to scan VCF (ie, variant call format) files for mutations in candidate genes, as we noted that ARIBA can miss detection of multiple nucleotide polymorphisms and complex single nucleotide polymorphisms. Resistance was predicted if a single determinant or multiple determinants for the same antibiotic were found; multiple determinants were not weighted or treated hierarchically.

For the **ARIBA software** see https://github.com/sanger-pathogens/ariba

For the **fastaq software** see https://github.com/sanger-pathogens/Fastaq

See **Online** for appendix 1

See **Online** for appendix 2

Assessing the diagnostic performance of genotypic predictions

Phenotypic categories were used as the reference standard and compared against genotypic predictions made using our curated database. Additionally, phenotypic categories were also compared with genotypic predictions made by LRE-Finder[9] version 1.0.0, ResFinder[5] version 4.1.10

(which includes PointFinder, version downloaded on March 3, 2022), and AMRFinderPlus[4] version 3.10.18 (database version 2021-12-21.1).[33] We classified isolates as the following: (1) true positive, if phenotypically resistant (or non-susceptible) with a known resistance-conferring genetic determinant or determinants detected in their genome; (2) true negative, if phenotypically susceptible in the absence of any genetic determinant; (3) false positive, if phenotypically susceptible in the presence of a known genetic determinant; and (4) false negative, if phenotypically resistant (or non-susceptible) but not carrying a known resistance-conferring genetic determinant in their genome.

Statistical analysis

We calculated commonly used metrics of diagnostic performance using the epi.tests function from the epiR R package version 2.0.43 on R version 4.1.0, including the following: negative predictive value (NPV), positive predictive value (PPV), sensitivity, and specificity. 95% CIs were calculated using Wilson's score. We additionally calculated the diagnostic performance of resistance and non-susceptibility genotypic predictions using only EUCAST clinical breakpoints or only CLSI breakpoints.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

The number of isolates, antibiotics tested, AST method used, and source of isolates for each study of available collections of *E faecium* are presented in table 1, and appendix 3 (tab 1) provides isolate-level metadata. To establish a database of genetic determinants of antibiotic resistance in *E faecium*, we curated 316 genetic determinants, consisting of single mutations (n=103), combinations of mutations (n=100), single acquired genes (n=82), and multiple acquired genes (ie, operons; n=27) for a total of 17 antibiotics—ie, the 15 antibiotics selected plus streptogramin A and streptogramin B, as resistance determinants to both streptogramins are needed to predict quinupristin–dalfopristin resistance. We encoded for cross-resistance across tetracyclines (ie, doxycycline and tetracycline). As individual genetic determinants can influence susceptibility to multiple antibiotics, susceptibility to each of the antibiotics was encoded as a separate relationship and defined by an effect classified as resistance (n=228), susceptibility (n=4), or undetermined (n=84) on the basis of the existence or scarcity of evidence in the scientific literature. Only the resistance category was used to make genotypic predictions; a summary of the genetic determinants of resistance is provided in table 2. Appendix 4 (tab 1) includes the complete database, including the genetic determinants with undetermined effect.

See Online for appendix 3

See Online for appendix 4

See Online for appendix 5

Table 3 summarises the diagnostic performance in terms of sensitivity, specificity, PPV, and NPV of using the curated database to determine phenotypic resistance. We calculated these diagnostic metrics to predict both phenotypic resistance in the 12 selected antibiotics (table 3) and phenotype non-susceptibility in five of these antibiotics (appendix 2 p 20), as described in the Methods. The accuracy of genotypic predictions for erythromycin, clindamycin, and kanamycin, not commonly used to treat enterococci infections, is presented in appendix 2 (pp 10–11, 25).

Resistance to ampicillin (>8 µg/mL MIC) was predicted with very high sensitivity (99·7%, 95% CI 99·5–99·9) and specificity (97·9%, 95·8–99·0; table 3). Ampicillin non-susceptibility (>4 µg/mL MIC) was predicted with the same specificity (97·9%, 95·8–99·0) and very high sensitivity too (99·6%, 99·3–99·8; appendix 2 p 20), due to the presence of ampicillin-resistant *pbp5* mutations in isolates with a MIC of 8 µg/mL (figure 1A). Ampicillin resistance was attributable to *pbp5* mutations only (appendix 2 p 12), and no isolate carried the *blaZ* operon. Resistance to ciprofloxacin (>4 µg/mL MIC) was predicted with very high sensitivity (98·0%, 96·4–98·9) and specificity (98·8%, 95·9–99·7), and was attributable to combinations of *gyrA* and *parC* mutations (table 3; figure 1B).

Resistance to vancomycin (>4 µg/mL MIC) was predicted with very high sensitivity (98·8%, 95% CI 98·3–99·2) and specificity (98·8%, 98·0–99·3; table 3). Of the 2309 vancomycin-resistant isolates, the *vanA* operon was detected in 2178 (94·3%) isolates, *vanB* operon in 96 (4·2%) isolates, and both operons in seven (0·3%) isolates. Because *vanA*-positive vancomycin-susceptible isolates have been described,[34] caused by the loss of specific genes in the *vanA* operon, we encoded resistance on the basis of the presence and integrity of all genes in the operon, not on the presence of *vanA* or *vanB* genes alone. We encoded for the full *vanA* operon but also for variants of the *vanA* operon without specific genes (table 2), as we found such variants in phenotypically resistant isolates (figure 1C; appendix 2 pp 13–14). If vancomycin resistance had been predicted on the basis of the detection of *vanA* or *vanB* genes alone, this decision would have led to a 5·5 percentage point drop in specificity (93·3%, 91·7–94·6) and no increase in sensitivity (98·7%, 98·2–99·1).

Resistance to teicoplanin (>2 µg/mL MIC) was predicted with very high sensitivity (98·9%, 95% CI 98·4–99·3) but slightly lower specificity (93·3%, 91·3–94·9; table 3). Specificity was only moderate due to 52 isolates carrying a *vanA* operon variant but being phenotypically susceptible to teicoplanin (figure 1D). The majority of these false positives (44 [85%] of 52) were phenotypically resistant to vancomycin (>4 µg/mL MIC), thus displaying what has been previously described as a VanB phenotype–*vanA* genotype.[35] Of these 44 isolates, 12 isolates from the same collection[28] were re-tested and found to have a heteroresistance phenotype (appendix 5 tab 1), as shown by colonies within the resistant zone on E-test strips (bioMérieux, Marcy l'Étoile, France).

Resistance to doxycycline and tetracycline was predicted with very high sensitivity (97·0% [95% CI 91·5–99·0] for doxycycline and 99·5% [99·1–99·7] for tetracycline) but low

	Isolates, n	Antibiotics tested and used	AST methods	Geographical origin (year)	Sources
Gouliouris et al (2021)[28]	1658	Ampicillin, vancomycin, teicoplanin, tetracycline, tigecycline, streptomycin, gentamicin, kanamycin, erythromycin, clindamycin, quinupristin–dalfopristin, linezolid	Vitek 2 P607 (bioMérieux, Marcy l'Étoile, France)	UK (2015)	Clinical
Raven et al (2016)[30]	495	Ampicillin, vancomycin, teicoplanin, ciprofloxacin, gentamicin, erythromycin, linezolid	Vitek 2 P607 (bioMérieux, Marcy l'Étoile, France) or ADM	UK (2001–11)	Clinical
Gouliouris et al (2019)[31]	398	Ampicillin, vancomycin, teicoplanin, tetracycline, tigecycline, streptomycin, gentamicin, kanamycin, erythromycin, clindamycin, quinupristin–dalfopristin, linezolid	Vitek 2 P607 (bioMérieux, Marcy l'Étoile, France)	UK (2014–15)	Wastewater
Zaheer et al (2020)[21]*	336	Ampicillin, vancomycin, teicoplanin, doxycycline, tigecycline, streptomycin, gentamicin, erythromycin, quinupristin–dalfopristin, linezolid	Disc diffusion	USA (2014–16)	Clinical, cattle, beef, water, and wastewater
Raven et al (2017)[29]	293	Ampicillin, vancomycin, teicoplanin, tetracycline, tigecycline, streptomycin, gentamicin, kanamycin, erythromycin, clindamycin, quinupristin–dalfopristin, linezolid	Vitek 2 P607 (bioMérieux, Marcy l'Étoile, France)	UK (2006–12)	Clinical
Gouliouris et al (2018)[27]	267	Ampicillin, vancomycin, teicoplanin, tetracycline, tigecycline, streptomycin, gentamicin, kanamycin, erythromycin, clindamycin, quinupristin–dalfopristin, linezolid	Vitek 2 P607 (bioMérieux, Marcy l'Étoile, France)	UK (2014–15)	Chicken, turkey, pig (both live animals and meat), and cattle
ICU study, Cambridge, UK (unpublished)†	243	Ampicillin, vancomycin, teicoplanin, tetracycline, tigecycline, streptomycin, gentamicin, kanamycin, erythromycin, clindamycin, quinupristin–dalfopristin, linezolid	Vitek 2 P607 (bioMérieux, Marcy l'Étoile, France)	UK (2016)	Clinical
Rushton-Green et al (2019)[23]*	116	Ampicillin, vancomycin, erythromycin	Broth microdilution (no device specified)	New Zealand (2000–03)	Clinical and poultry
Wang et al (2018)[24]*	106	Vancomycin, daptomycin	Broth microdilution (MicroScan 96Plus; location of manufacture not specified)	USA (2009–13)	Clinical
Tyson et al (2018)[11]*	100	Vancomycin, daptomycin, ciprofloxacin, tetracycline, tigecycline, streptomycin, gentamicin, kanamycin, erythromycin, quinupristin–dalfopristin, linezolid	Broth microdilution (CMV3AGPF plates; TREK Diagnostic Systems, Independence, OH, USA)	USA (2014)	Cattle, chicken, turkey (both live animals and meat), and pork
Roberts et al (2022)[32]	81	Ampicillin, vancomycin, teicoplanin, tetracycline, tigecycline, streptomycin, gentamicin, kanamycin, erythromycin, clindamycin, quinupristin–dalfopristin, linezolid	Vitek 2 P607 (bioMérieux, Marcy l'Étoile, France)	Viet Nam (2017–18)	Clinical
Bortolaia et al (2020)[12]*	50	Ampicillin, vancomycin, ciprofloxacin, tetracycline, gentamicin, erythromycin, linezolid	Broth microdilution (no device specified)	Denmark (not reported)	Clinical
Zankari et al (2013)[3]*	49	Ampicillin, vancomycin, teicoplanin, ciprofloxacin, tetracycline, tigecycline, streptomycin, gentamicin, kanamycin, erythromycin, linezolid	Broth microdilution (no device specified)	Denmark (2011)	Pigs
Zhou et al (2018)[15]*	37	Vancomycin	Vitek 2 (location of manufacture not specified)	Netherlands (2014)	Clinical
Hasman et al (2019)[9]*	36	Linezolid	Broth microdilution (no device specified)	Denmark (2015–18)	Clinical
Roach et al (2015)[16]*	35	Ampicillin, vancomycin, daptomycin, ciprofloxacin, tetracycline, tigecycline, erythromycin, linezolid	Broth microdilution (Sensititre; TREK Diagnostic Systems; location of manufacture not specified)	USA (2012–13)	Clinical
Prater et al (2019)[18]*	29	Ampicillin, vancomycin, daptomycin, erythromycin	Broth microdilution (no device specified)	Not reported	Clinical
Kessel et al (2021)[19]*	28	Tigecycline	Vitek 2 (bioMérieux, Nuertingen, Germany)	Germany (2014–17)	Clinical

(Table 1 continues on next page)

	Isolates, n	Antibiotics tested and used	AST methods	Geographical origin (year)	Sources
(Continued from previous page)					
Kwak et al (2020)[20]*	25	Ampicillin, vancomycin, doxycycline, linezolid	Disc diffusion	USA and Canada (2014–15)	Clinical
Overall	4382	Ampicillin, vancomycin, teicoplanin, daptomycin, ciprofloxacin, tetracycline, doxycycline, tigecycline, streptomycin, gentamicin, kanamycin, erythromycin, clindamycin, quinupristin–dalfopristin, linezolid	Vitek 2, disc diffusion, broth microdilution, broth microdilution with Sensititre, and broth microdilution with MicroScan 96Plus	UK, Denmark, Netherlands, Germany, Austria, USA, Canada, Australia, New Zealand, Viet Nam, and Pakistan	Clinical, water, wastewater, chicken, turkey, and pig, cattle (samples obtained from both live animals and meat)

Characteristics and number of *Enterococcus faecium* isolates of the collections used in this study. Collections are ordered by the number of available *E faecium* isolates. Isolate-level metadata, including genome accession numbers and AST phenotypic metadata can be found in appendix 3 (tab 1). Only antibiotics with at least 50 susceptible and 50 resistant isolates across studies were considered and presented in this table. AST=antibiotic susceptibility testing. ADM=agar dilution method. ICU=intensive-care unit. *Publicly available. †Isolates obtained from the ICU of Addenbrookes Hospital, Cambridge, UK, by HJW, FAK, and MET.

Table 1: **Characteristics of *Enterococcus faecium* collections used in this study**

	Single acquired genes		Multiple acquired genes		Single mutations		Multiple mutations	
	n	Gene names	n	Operon or gene names	n	Gene names	n	Gene names
Ampicillin	0	NA	1	*blaZ* operon	10	*pbp5* (n=10)	7	*pbp5* (n=7)
Vancomycin	0	NA	11	*vanA* operon variants* and *vanB* operon variants†	0	NA	0	NA
Teicoplanin	0	NA	5	*vanA* operon variants*	0	NA	0	NA
Daptomycin	0	NA	0	NA	14	*cls* (n=7), *liaF* (n=2), *liaR* (n=2), *liaS* (n=3)	28	*lia* and *cls* (n=14), *lia* genes (n=3), and others (n=11)‡
Ciprofloxacin	0	NA	0	NA	15	*gyrA* (n=12), *parC* (n=3)	30	*gyrA*, *gyrB*, and *parC* (n=2), *gyrA* and *parC* (n=28)
Tetracycline	9	*tetL*, *tetM*, *tetO*, *tetS*, *tetWNW*, *tetK*, *tetT*, *tetW*, *tet40*	0	NA	0	NA	0	NA
Tigecycline	1	*tetX1*	0	NA	7	*rpsJ* (n=7)	8	*rpsJ* (n=8)
Streptomycin§	5	*ANT(3′′)-Ia*, *aad(6)*, *ANT(6)-Ia*, *ANT(6)-Ia-pAML0157*, *APH(2′′)-Ie*	0	NA	0	NA	0	NA
Gentamicin§	6	*AAC(6′)-Ie-APH(2′′)-Ia*, *APH(2′′)-IIa*, *APH(2′′)-IIIa*, *APH(2′′)-IVa*, *APH(6)-Ic*	0	NA	0	NA	0	NA
Kanamycin§	6	*AAC(6′)-Ie-APH(2′′)-Ia*, *APH(2′′)-IIa*, *APH(2′′)-IIIa*, *APH(2′′)-IVa*, *APH(3′)-IIIa*, *ANT(4′)-Ia*	0	NA	0	NA	0	NA
Erythromycin	9	*ermB*, *ermB-E7067*, *ermA*, *ermA-AF002716*, *ermA-EU348758*, *ermA-KT862775*, *ermT*, *ermC*, *mefA*	0	NA	0	NA	0	NA
Clindamycin	14	*ermB*, *ermB-E7067*, *ermA*, *ermA-AF002716*, *ermA-EU348758*, *ermA-KT862775*, *lsaA*, *lsaB*, *lsaC*, *lsaE*, *lnuA*, *lnuB*, *lnuC*, *lnuF*	0	NA	1	*eatA* (n=1)	0	NA
Streptogramin A¶	7	*lsaA*, *lsaB*, *lsaC*, *vgaD*, *vatH*, *vatD*, *vatE*	0	NA	1	*eatA* (n=1)	0	NA
Streptogramin B¶	6	*lsaA*, *lsaB*, *lsaC*, *vgbA*, *ermB*, *ermB-E7067*	0	NA	0	NA	0	NA
Quinupristin–dalfopristin	0	NA	1	*vatD* and *vgbA* genes	0	NA	0	NA
Linezolid	4	*cfrA*, *cfrB*, *optrA*, *poxtA*	0	NA	20	*23SrRNA* (n=15), *rplC* (n=1), *rplD* (n=3), *rplV* (n=1)	2	*23SrRNA* (n=1), *23SrRNA* and *rplC* (n=1)

This table summarises 228 genetic determinants included in the curated database and predictive of antibiotic resistance in *Enterococcus faecium*. See appendix 4 (tabs 1–4) for the complete database. Cross-resistance across fluoroquinolones (ie, ciprofloxacin) and across tetracyclines (ie, doxycycline and tetracycline) was assumed, hence the absence of a row for doxycycline. NA=not applicable. **vanA* operon variants include: the complete *vanA* operon, *vanAΔvanS*$_A$, *vanAΔvanYA*, *vanAΔvanZA*, and *vanAΔvanYAΔvanZA*. †*vanB* operon variants include: the complete *vanB* operon, *vanB*-pE7948 variant (isolated from *E faecium* isolate E7948 plasmid 3 [NZ_LR135359·1]), *vanB*-pE7948Δ*vanWB*-pE7948, *vanB*-pE7948Δ*vanB*-pE7948::*vanB*, and *vanBΔvanRB*::*vanRB*-pE7948. ‡Other rare combinations of mutations associated with daptomycin resistance. §Acquired genes were used to predict high-level aminoglycoside resistance. ¶Quinupristin–dalfopristin resistance was predicted if isolates carried determinants for both streptogramin A and streptogramin B.

Table 2: **Summary of antibiotic resistance genetic determinants in *Enterococcus faecium* included in the curated database**

specificity (81·0% [75·3–85·6] and 71·0% [66·8–74·8]; table 3). Resistance to tetracycline (n=1869) was caused by single and combinations of acquired *tet* genes (figure 1E). Previous studies have also reported very high false positive rates for tetracyclines in *E faecium*,[10,11] caused by the detection of *tet* genes in phenotypically susceptible isolates. The low specificity observed could be attributable to isolates with an incorrect phenotypic AST result or the presence of

	Resistant isolates	Susceptible isolates	True positive	False negative	True negative	False positive	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
Ampicillin (n=3132)	2792 (89·1%)	340 (10·9%)	2785	7	333	7	99·7% (99·5–99·9)	97·9% (95·8–99·0)	99·7% (99·5–99·9)	97·9% (95·8–99·0)
Vancomycin (n=3398)	2309 (68·0%)	1089 (32·0%)	2281	28	1076	13	98·8% (98·3–99·2)	98·8% (98·0–99·3)	99·4% (99·0–99·7)	97·5% (96·4–98·2)
Teicoplanin (n=2760)	1982 (71·8%)	778 (28·2%)	1961	21	726	52	98·9% (98·4–99·3)	93·3% (91·3–94·9)	97·4% (96·6–98·0)	97·2% (95·7–98·2)
Daptomycin (n=220)	121 (55·0%)	99 (45·0%)	89	32	93	6	73·6% (65·1–80·6)	93·9% (87·4–97·2)	93·7% (86·9–97·1)	74·4% (66·1–81·2)
Ciprofloxacin (n=673)	501 (74·4%)	172 (25·6%)	491	10	170	2	98·0% (96·4–98·9)	98·8% (95·9–99·7)	99·6% (98·5–99·9)	94·4% (90·1–97·0)
Tetracycline (n=2358)	1869 (79·3%)	489 (20·7%)	1860	9	347	142	99·5% (99·1–99·7)	71·0% (66·8–74·8)	92·9% (91·7–94·0)	97·5% (95·3–98·7)
Doxycycline (n=320)	99 (30·9%)	221 (69·1%)	96	3	179	42	97·0% (91·5–99·0)	81·0% (75·3–85·6)	69·6% (61·4–76·6)	98·4% (95·3–99·4)
Tigecycline (n=2412)	60 (2·5%)	2352 (97·5%)	23	37	2346	6	38·3% (27·1–51·0)	99·7% (99·4–99·9)	79·3% (61·6–90·2)	98·4% (97·9–98·9)
Streptomycin* (n=2635)	1152 (43·7%)	1483 (56·3%)	1125	27	1314	169	97·7% (96·6–98·4)	88·6% (86·9–90·1)	86·9% (85·0–88·7)	98·0% (97·1–98·6)
Gentamicin* (n=3125)	1610 (51·5%)	1515 (48·5%)	1558	52	1246	269	96·8% (95·8–97·5)	82·2% (80·2–84·1)	85·3% (83·6–86·8)	96·0% (94·8–96·9)
Quinupristin–dalfopristin (n=1864)	314 (16·8%)	1550 (83·2%)	280	34	1465	85	89·2% (85·2–92·1)	94·5% (93·3–95·5)	76·7% (72·1–80·8)	97·7% (96·8–98·4)
Linezolid (n=2957)	65 (2·2%)	2892 (97·8%)	38	27	2843	49	58·5% (46·3–69·6)	98·3% (97·8–98·7)	43·7% (33·7–54·1)	99·1% (98·6–99·4)
Re-test†	38 (1·3%)	2919 (98·7%)	38	0	2870	49	100·0% (90·8–100·0)	98·3% (97·8–98·7)	43·7% (33·7–54·1)	100·0% (99·9–100·0)

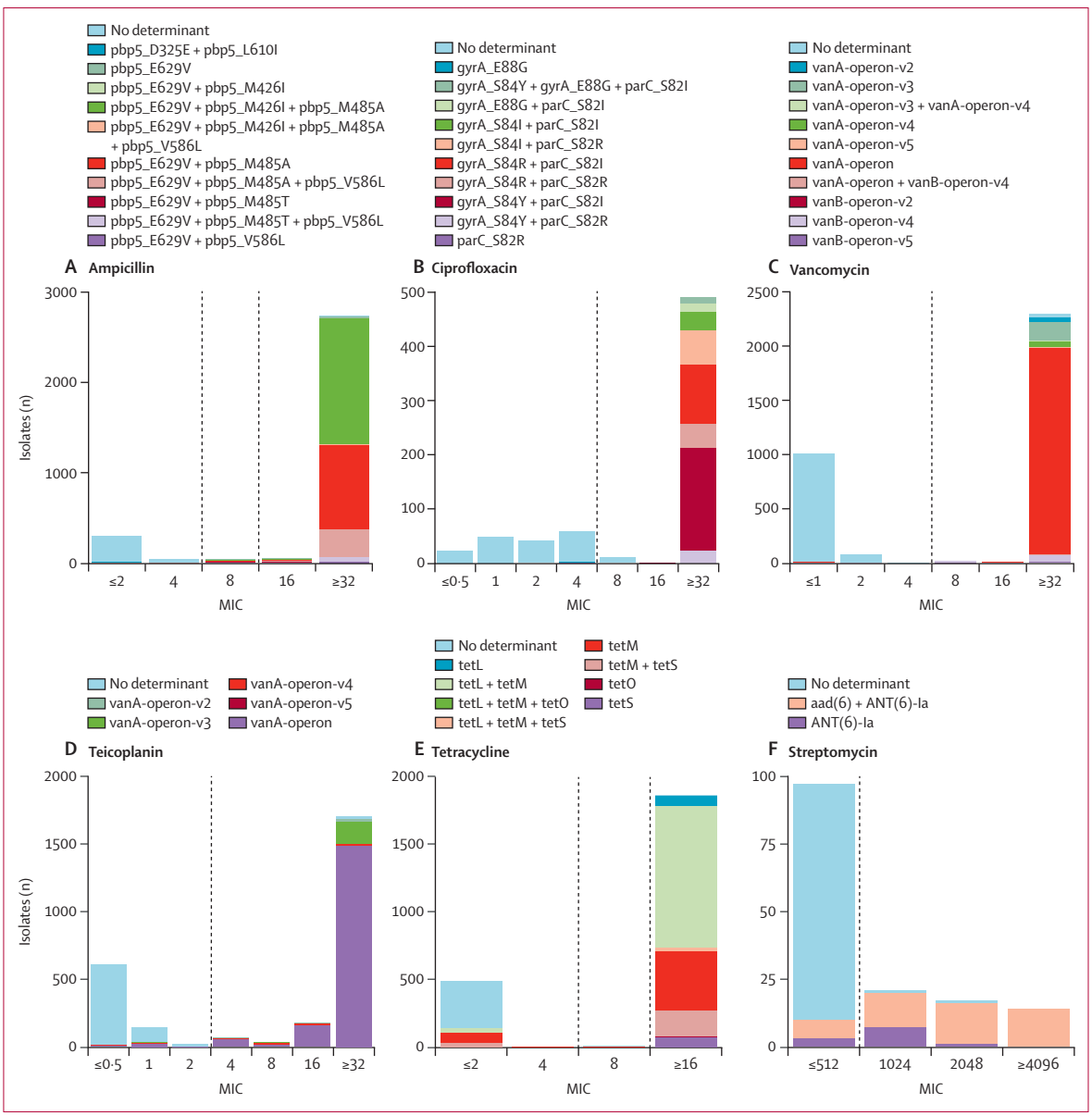
Data are n (%) or n, unless otherwise specified. Metrics of diagnostic performance when using the curated database to predict phenotypic antibiotic resistance. Here the intermediate susceptibility category is ignored—ie, isolates with intermediate susceptibility are treated as not tested. See appendix 2 (p 19) for the diagnostic performance achieved when predicting phenotypic non-susceptibility—ie, when treating the intermediate and resistance categories as non-susceptible. See appendix 2 (pp 17–18) for minimum inhibitory concentration and zone diameter breakpoints used in this study. A true positive was defined as a phenotypically resistant isolate carrying an antimicrobial resistance genetic determinant; a false negative was defined as a phenotypically resistant isolate not carrying any antimicrobial resistance genetic determinant; a true negative was defined as a phenotypically susceptible isolate not carrying any antimicrobial resistance genetic determinant; and a false positive was defined as a phenotypically susceptible isolate carrying an antimicrobial resistance genetic determinant. PPV=positive predictive value. NPV=negative predictive value. *High-level resistance to aminoglycoside. †Accuracy of genotypic predictions for linezolid after re-testing 27 false negatives that were susceptible according to E-test (bioMérieux, Marcy l'Etoile, France).

Table 3: **Accuracy of genotypic predictions of phenotypic resistance of *Enterococcus faecium* to 12 antibiotics when using the curated database**

silenced AMR-associated genes. To clarify the reason for the low specificity, we plotted the distribution of false positives on the phylogenetic tree (appendix 2 p 14) and re-tested a representative subset (re-testing criteria are specified in appendix 2 pp 6–7). As the high degree of clustering of false positives might be indicative of silenced *tet* genes, and a genetic basis for this silencing, we re-tested five false positives clustered in large clades of false positives (presumably more likely to represent cases of AMR-associated gene silencing than non-clustered cases), and five sporadic false positives surrounded by true positives of the same or neighbouring phylogenetic clade (presumably more likely to represent a phenotypic testing error than clustered false positives). As hypothesised, all five clustered false positives had a susceptible phenotype according to E-test (appendix 5 tab 1), indicating that their *tet* genes are silenced, whereas four of the five sporadic false positives were resistant according to E-test, indicating an erroneous original Vitek 2 result. Altogether, these results point to phenotypic testing errors and silenced *tet* genes as major sources of false positive genotypic predictions for tetracyclines.

High-level resistance to aminoglycosides (MIC >512 µg/mL for streptomycin and >128 µg/mL for gentamicin) was predicted with very high sensitivity, with 97·7% (95% CI 96·6–98·4) for streptomycin and 96·8% (86·9–90·1) for gentamicin (table 3). False positive rates were high, due to the presence of aminoglycoside-resistance genes in isolates with MICs below the high-level resistance breakpoint (figure 1F, 1G), resulting in genotypic predictions of low specificity (88·6% [86·9–90·1] for streptomycin and 82·2% [80·2–84·1] for gentamicin). The low specificity observed could be attributable to isolates with an elevated MIC below the high-level resistance breakpoint, a wrong phenotypic AST result, or the presence of silenced AMR-associated genes. To investigate the reason for the low specificity, we focused on isolates with available gentamicin MICs (n=534), although aminoglycoside MICs were not available for most isolates in our dataset. Of the 69 false positives with available gentamicin MICs, 33 had MICs above the epidemiological cutoff value (>32 µg/mL) and 36 had MICs below this value (appendix 2 p 15); the 36 isolates with MICs of 32 µg/mL or less carried putatively silenced aminoglycoside-resistance genes. To provide evidence for the presence of silenced aminoglycoside-resistance genes, we re-tested ten false positives with Vitek MICs below the epidemiological cutoff value (≤32 µg/mL) using E-test, including five clustered and five sporadic cases. The E-test results (appendix 5 tab 1) showed a mixture of gentamicin resistant (>32 µg/mL; n=7) and susceptible (≤32 µg/mL; n=3) isolates, with confirmed susceptibility results indicating the presence of silenced aminoglycoside-resistance genes.

The sensitivity and specificity of genotypic resistance to quinupristin–dalfopristin were reasonably high (89·2% [95% CI 85·2–92·1] sensitivity and 94·5% [93·3–95·5] specificity), although lower than that achieved for other antibiotics (table 3). Phenotypic resistance (>4 µg/mL MIC; n=314) was explained by the coincident detection of the *eatA* Thr450Ile mutation, conferring resistance to dalfopristin, and the *ermB* gene, conferring resistance to quinupristin (280 [89·2%] true positives out of 314 resistant isolates; figure 1H; table 3). Eight of these 280 isolates also carried *vatE*. Phenotypic non-susceptibility (>1 µg/mL MIC; n=788) was predicted with lower sensitivity (81·1%, 78·2–83·7) and the same specificity (appendix 2 p 20).



(Figure 1 continues on next page)

The sensitivities of genotypic resistance predictions to the last-line antibiotics daptomycin, tigecycline, and linezolid were the lowest (73·6% [95% CI 65·1–80·6], 38·3% [27·1–51·0], and 58·5% [46·3–69·6], respectively) compared with all other antibiotics investigated (table 3; figure 1I–K), although specificities were high (93·9% [87·4–97·2], 99·7% [99·4–99·9], and 98·3% [97·8–98·7]). We detected an unexpectedly high number of false negatives for linezolid (n=27), whose phenotypic resistance should have been explained by known resistance mechanisms. We therefore re-tested all false negatives (appendix p 5 tab 1) using E-test, which displayed a susceptible linezolid phenotype (≤4 μg/mL MIC). When considering these new phenotypic results, the sensitivity of genotypic predictions for linezolid became 100·0% (90·8–100·0; table 3).

We additionally compared the effect of using either CLSI or EUCAST clinical breakpoints (appendix 2 p 13) on the diagnostic accuracy of genotypic predictions, and found that the use of EUCAST clinical breakpoints resulted in better phenotype–genotype concordance for ampicillin, ciprofloxacin, and teicoplanin resistance than did CLSI breakpoints (appendix 2 pp 9–10).

As a comparison, we also determined antibiotic resistance using the latest versions of AMRFinderPlus[4] (appendix 2 p 21), ResFinder[5] (appendix 2 pp 22–23), and LRE-Finder[9] for linezolid resistance (appendix 2 p 23). Relationships of

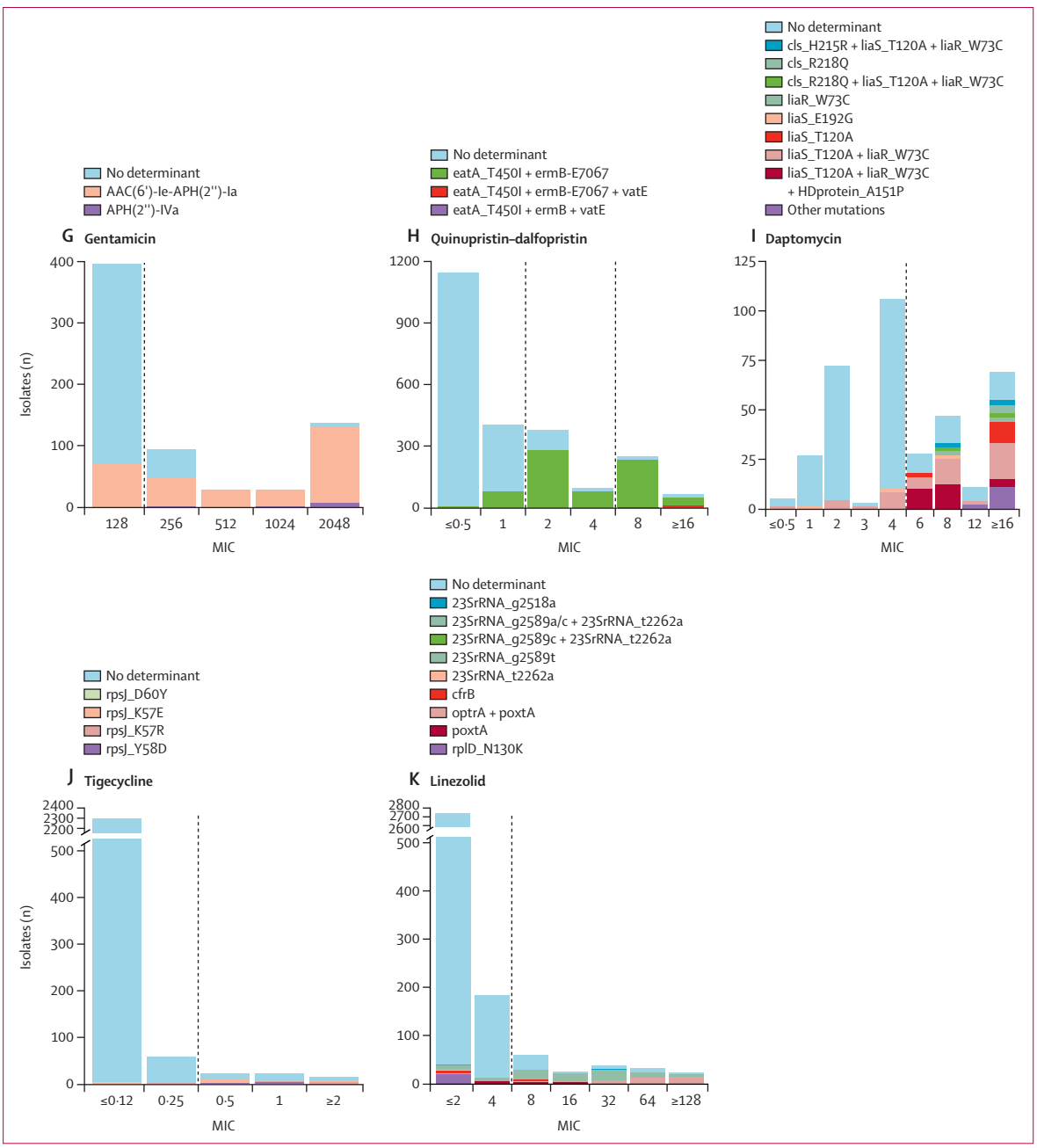


Figure 1: **MIC distributions for isolates with AMR determinants included in the curated database**
These bar plots show the number of isolates with different MICs per antibiotic. AMR determinants are referred to by their identifiers listed in the curated database (id column in appendix 4 tab 1). The left-most MIC category includes all isolates with MIC equal to or lower than the MIC value specified. The right-most MIC category includes all isolates with MIC equal to or higher than the MIC value specified. The distribution of AMR genetic determinants within each bar (MIC category) is shown in different colours. Vertical dotted lines indicate the MIC breakpoints used in this study (appendix 2 p 17). For doxycycline, most antibiotic susceptibility testing data available were obtained with the disc diffusion method, so no MIC data were available for plotting. AMR=antimicrobial resistance. MIC=minimum inhibitory concentration.

cross-resistance were assumed among antibiotics of the same antibiotic class when individual antibiotics were not explicitly predicted by an AMR database (appendix 2 p 24). These tools produced similar sensitivities and specificities (figure 2) for resistance to ciprofloxacin (ResFinder and AMRFinderPlus) and linezolid (LRE-Finder), and high-level resistance to aminoglycosides (AMRFinderPlus). Resistance to vancomycin, teicoplanin, and tetracycline was predicted with the same sensitivity as the curated database, but always at lower specificity. For glycopeptides, lower specificity was caused by truncated variants of the *vanA* ($vanA\Delta vanR_A\Delta vanS_A$) and *vanB* ($vanB\Delta vanR_B\Delta vanS_B$)

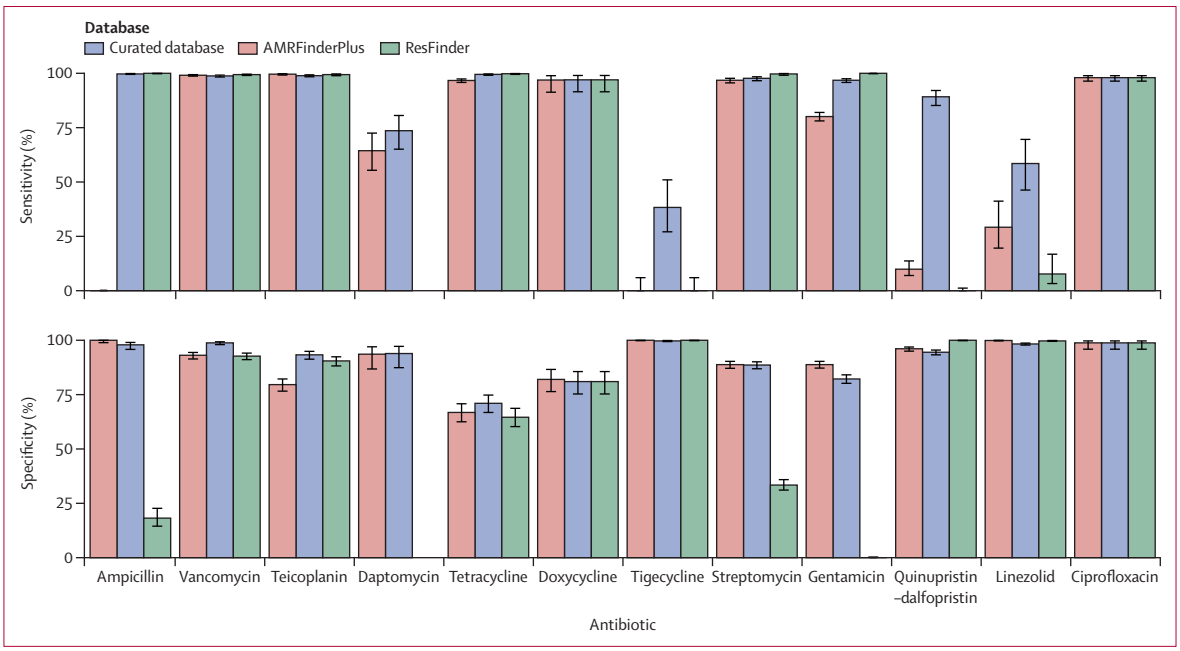


Figure 2: **Diagnostic accuracy of genotypic antibiotic resistance predictions of the curated database compared with other databases**
This figure shows the sensitivity and specificity of genotypic antibiotic resistance predictions when using our curated database compared with the values obtained by AMRFinderPlus and ResFinder. Error bars represent the 95% CIs. See Methods for details on database versions used.

operons detected in phenotypically susceptible isolates. For tetracycline, the lower specificity is explained by the detection of a truncated variant of *tetM* (a frameshift mutation at Lys261fs) in phenotypically susceptible isolates. Very low specificities were obtained for ampicillin, clindamycin, and erythromycin, preventing the use any of these tools to predict resistance for these antibiotics. For ampicillin, the inclusion of *pbp5* mutations in the ResFinder database commonly detected in phenotypically susceptible isolates resulted in very poor specificity. Very high specificity in predictions, similar to that obtained by the curated database, was obtained for daptomycin (AMRFinderPlus), tigecycline (all three tools), linezolid (ResFinder and AMRFinderPlus), and quinupristin–dalfopristin (ResFinder and AMRFinderPlus), albeit with lower sensitivities than those obtained by the curated database. Lower sensitivity was caused by the absence of known resistance-conferring determinants in these databases. The diagnostic performance of LRE-Finder to predict linezolid resistance (100·0% [95% CI 94·4–100·0] sensitivity and 99·0% [98·6–99·3] specificity; appendix 2 p 23) was very similar to that obtained by the curated database (table 3).

We additionally used a collection of 382 *E faecium* isolates published in 2022[10] as an independent dataset to validate the performance of our curated database. Despite the smaller size of this dataset, we achieved better prediction accuracy (appendix 2 p 25) for ampicillin (100·0% [95% CI 98·8–100·0] sensitivity and 97·4% [86·8–99·5] specificity), vancomycin (99·0% [97·1–99·7] and 98·5% [92·0–99·7]), and high-level gentamicin resistance (93·3% [78·7–98·1] and 57·9% [52·5–63·0]), and the same accuracy for ciprofloxacin (100·0% [98·8–100·0] and 100·0% [92·0–100·0]), tetracycline (98·9% [96·8–99·6] and 87·6% [79·2–92·9]), and doxycycline (99·5% [97·2–99·9] and 72·2% [63·1–79·8]), compared with AMRFinderPlus (appendix 2 p 26) and ResFinder (appendix 2 p 26). Very low resistance rates for linezolid (only two resistant isolates), and scarcity of testing for tigecycline and daptomycin susceptibility, precluded validation of genotypic resistance detection for these drugs.

Discussion

To our knowledge, this work represents the largest effort conducted to date to evaluate the accuracy of antibiotic resistance predictions from *E faecium* genomes. We curated a database of 228 genetic determinants known to cause resistance to 12 different, clinically relevant antibiotics in *E faecium*, and evaluated the diagnostic performance of this catalogue in a dataset of 4382 *E faecium* genomes. We showed that very accurate resistance predictions could be made for ampicillin, ciprofloxacin, vancomycin, linezolid, teicoplanin, and quinupristin–dalfopristin. The specificity of genotypic predictions needs improvement for tetracyclines and high-level resistance to aminoglycosides, and sensitivity needs improvement for tigecycline and daptomycin.

Although mutations in penicillin-binding protein 5 (Pbp5) have long been considered the major contributor to ampicillin resistance in *E faecium*, knowledge of what specific amino acid changes in this protein contribute to resistance has been scarce.[3,11] This lack of knowledge was due to the scarcity of whole-genome sequences, which made it only possible to compare the *pbp5*-resistance allele carried by clade A1 isolates with the *pbp5*-susceptible allele carried by

clade B isolates, which differ by 21 amino acids.[36] Here, the availability of diverse clade A2 isolates with different ampicillin susceptibility (appendix 2 p 12) allowed us to differentiate the key amino acid changes contributing to ampicillin resistance.

Resistance to glycopeptides was predicted with very high sensitivity and specificity by encoding and accounting for all the different functional variants of the *vanA* and *vanB* operons. Predicting resistance by detecting the *vanA* or *vanB* genes alone would have resulted in drops in specificity, due to the presence of truncated variants of these operons in phenotypically susceptible isolates (appendix 2 p 13). Otherwise, predicting resistance by detecting the full *van* operons would have resulted in lower sensitivity, as operon variants missing specific genes are found in phenotypically resistant isolates (appendix 2 p 13). These findings are consistent with the non-essential roles of *vanZ* and *vanY*, and the loss of *vanS*, which leads to constitutive expression but does not compromise the expression of vancomycin resistance. Further work should assess the contribution of less common *van* operons (eg, *vanC*, *vanD*, *vanE*, *vanG*, *vanL*, *vanM*, and *vanN*) to glycopeptide resistance in *E faecium*.

Commonly used AMR databases produced worse genotypic predictions than our curated database for many antibiotics. The detection of non-functional *van* operon variants, truncated variants of genes (eg, a frameshift mutation in *tetM*), non-conferring resistance mutations (eg, in *pbp5*) and genes (eg, *msrC*), and incorrectly encoded gene–resistance relationships (eg, *ermT*) were all reported causes of low specificity. These results emphasise the requirement of accounting not only for the presence but also for the integrity of antibiotic resistance genes (ie, fully assembled without frameshifts or truncations) when predicting resistance, and the importance of continuous and improved curation of existing databases.

Both our curated database and other tools achieved low specificity in the prediction of resistance to tetracycline and high-level resistance to aminoglycosides. In addition to erroneous phenotypic AST results being the source of these discrepancies, we showed that clustering of false positives in the phylogeny was common and indicated the presence of silenced AMR-associated genes after phenotypic re-testing. This phenomenon could represent what has previously been termed as silent acquired AMR.[37] Further research is needed to quantify the prevalence and genetic basis of this phenomenon in *E faecium*.

The absence of known resistance-conferring determinants in these databases explained the lower sensitivity obtained for tigecycline and daptomycin. Although the curated database achieved better sensitivities than other databases, further research is needed to fully characterise the genetic basis of resistance to these antibiotics. Providing further evidence for a causative link between the presence of a new AMR-associated gene or mutation with phenotypic resistance will require experimental validation. This validation is clinically important, as these last-resort antibiotics are used for the treatment of vancomycin-resistant *E faecium* infection.

Our study has several limitations. We made use of a diverse and opportunistic dataset of *E faecium* isolates with phenotypic AST determined using different methods. The prediction accuracy metrics calculated here should be validated in a larger independent collection, with isolates tested against the antibiotics investigated here, ideally with a single reference AST method (eg, broth microdilution). We limited our approach to make binary predictions of antibiotic susceptibility. However, antibiotic resistance is a quantitative trait, as shown by a wide range of MICs for some antibiotics (figure 1). Further work is needed to determine antibiotic resistance quantitatively from genomic data, particularly for complex and polygenic phenotypes such as resistance to daptomycin, and to discriminate between the intermediate and resistance categories to ultimately inform likelihood of therapeutic success. With the exception of linezolid, we did not re-test all instances of genotype–phenotype discrepancies, but instead re-tested a selection of isolates to explore the plausible cause of such discrepancies. Similarly, we did not systematically assess the contribution of heteroresistance (ie, the presence of both resistant and susceptible subpopulations) as a source of discrepancies, which can result in the selection of isolates for sequencing and AST with different underlying susceptibilities. Despite the large dataset used, the number of tested isolates was modest (about 200–400 isolates) for daptomycin and doxycycline, and the proportion of resistant isolates was low for tigecycline (2·5%) and linezolid (1·3%). The low prevalence of resistance to these antibiotics led to low PPVs and very wide 95% CIs around their estimates. Larger and more resistant *E faecium* populations are needed to obtain more accurate accuracy estimates for these antibiotics.

In conclusion, the accuracy of genotypic predictions of our curated database was equivalent or enhanced compared with existing databases. We additionally implemented these genotypic predictions on Pathogenwatch, a web-based tool that already supports the detection of AMR from the genomes of a variety of microbial pathogens, and now for *E faecium* too. A complete characterisation of the genetic basis of non-susceptibility to last-line antibiotics daptomycin and tigecycline, and the mechanisms mediating antibiotic resistance silencing, are needed to close the remaining sensitivity and specificity gaps in genotypic predictions. Given the increasingly routine use of whole-genome sequencing in clinical and public health microbiology laboratories, our study will facilitate the adoption of whole-genome sequencing as a tool for the diagnosis and surveillance of AMR in *E faecium*.

For more on **Pathogenwatch** see https://pathogen.watch/

Contributors

FC and TG designed the study with input from all other authors. FC undertook the bioinformatic analyses with contributions from TG. KER, CL, BB, EMH, CSH, NMB, MAH, JP, MET, and SJP all contributed to the generation of the genomic and antibiotic susceptibility data used in this study. HJW, FAK, and LWR contributed specifically to the laboratory work and genomic analyses for studies conducted in intensive-care units (ICUs) in Addenbrooke's Hospital, Cambridge, UK, and in Viet Nam, which provided data for this study. LTH, NTH, and NVT contributed to generating the data in the Viet Nam ICU study. BB conducted the AST work needed to

re-test isolates with incongruent genotypic predictions. CAY implemented genotypic predictions on Pathogenwatch. FC, TG, and SJP wrote the first draft of the manuscript. SJP and JP supervised the study. All authors had full access to all the data in the study and accept responsibility for the decision to submit for publication. All authors read, contributed to, and approved the final manuscript. FC and TG verified the underlying data of the study.

Declaration of interests

SJP and JP are consultants to Next Gen Diagnostics. SJP is also a consultant for Specific Technologies. All other authors declare no competing interests.

Data sharing

The whole-genome sequences of the strain collections used in this study are available on European Nucleotide Archive under the accessions listed in appendix 3 (tab 1), which also includes the antibiotic susceptibility phenotypes (appendix 3 tab 1) and genotypes (appendix 3 tabs 2–5). Appendix 4 (tabs 1–4) includes the curated database of antibiotic resistance genetic determinants in *E faecium*; appendix 5 (tab 1) the antibiotic susceptibility phenotypes of re-tested *E faecium* isolates; and appendix 1 the genome accessions, and antibiotic susceptibility phenotypes (appendix 1 tab 1) and genotypes (appendix 1 tabs 2–5) of all isolates in the validation dataset. All scripts necessary to run the described analyses are available on GitHub (https://github.com/francesccoll/amr_efm_wgs).

Acknowledgments

This publication presents independent research supported by Wellcome grants 201344/Z/16/Z and 204928/Z/16/Z awarded to FC. This publication was also supported by the Health Innovation Challenge Fund (WT098600, HICF-T5–342), a parallel funding partnership between the UK Department of Health and Wellcome Trust. We acknowledge the British Society for Antimicrobial Chemotherapy as a source of *E faecium* isolates in this study. The Cambridge ICU study that provided data for this study was supported by a Clinician Scientist Fellowship to MET, funded by the Academy of Medical Sciences and the Health Foundation, and by the UK National Institute for Health and Care Research Cambridge Biomedical Research Centre. The Viet Nam ICU study that provided data for this study was funded by the Medical Research Council Newton Fund (grant reference MR/N029399/1) and the Vietnamese Ministry of Science and Technology (grant reference HNQT/SPDP/04·16). CL was supported by a Wellcome Trust Sir Henry Wellcome Postdoctoral Fellowship (110243/Z/15/Z). This study has also been funded by a Research Grant 2020 of the European Society of Clinical Microbiology and Infectious Diseases awarded to KER. LWR was funded by an EBI–Sanger Postdoctoral (ESPOD) fellowship from the European Molecular Biology Laboratory. This research was supported by the NIHR Cambridge Biomedical Research Centre (NIHR203312). The views expressed in this publication are those of the authors and not necessarily those of the funders.

References

1 Murray CJ, Ikuta KS, Sharara F, et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet* 2022; **399:** 629–55.

2 Tacconelli E, Carrara E, Savoldi A, et al. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet Infect Dis* 2018; **18:** 318–27.

3 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:** 771–77.

4 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.

5 Florensa AF, Kaas RS, Clausen PTLC, Aytan-Aktug D, Aarestrup FM. ResFinder—an open online resource for identification of antimicrobial resistance genes in next-generation sequencing data and prediction of phenotypes from genotypes. *Microb Genom* 2022; **8:** 1–10.

6 Allix-Béguec C, Arandjelovic I, Bi L, et al. Prediction of susceptibility to first-line tuberculosis drugs by DNA sequencing. *N Engl J Med* 2018; **379:** 1403–15.

7 Bradley P, Gordon NC, Walker TM, et al. Rapid antibiotic-resistance predictions from genome sequence data for *Staphylococcus aureus* and *Mycobacterium tuberculosis*. *Nat Commun* 2015; **6:** 10063.

8 Gladstone RA, Lo SW, Lees JA, et al. International genomic definition of pneumococcal lineages, to contextualise disease, antibiotic resistance and vaccine impact. *EBioMedicine* 2019; **43:** 338–46.

9 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:** 1473–76.

10 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:** e0119621.

11 Tyson GH, Sabo JL, Rice-Trujillo C, Hernandez J, McDermott PF. Whole-genome sequencing based characterization of antimicrobial resistance in *Enterococcus*. *Pathog Dis* 2018; **76:** 1–5.

12 Bortolaia V, Kaas RS, Ruppe E, et al. ResFinder 4·0 for predictions of phenotypes from genotypes. *J Antimicrob Chemother* 2020; **75:** 3491–500.

13 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* 2022; **42:** 67–76.

14 Manson AL, Van Tyne D, Straub TJ, et al. Chicken meat-associated enterococci: influence of agricultural antibiotic use and connection to the clinic. *Appl Environ Microbiol* 2019; **85:** 1–14.

15 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.

16 Roach DJ, Burton JN, Lee C, et al. A year of infection in the intensive care unit: prospective whole genome sequencing of bacterial clinical isolates reveals cryptic transmissions and novel microbiota. *PLoS Genet* 2015; **11:** e1005413.

17 Douglas AP, Marshall C, Baines SL, et al. Utilizing genomic analyses to investigate the first outbreak of vanA vancomycin-resistant *Enterococcus* in Australia with emergence of daptomycin non-susceptibility. *J Med Microbiol* 2019; **68:** 303–08.

18 Prater AG, Mehta HH, Kosgei AJ, et al. Environment shapes the accessible daptomycin resistance mechanisms in *Enterococcus faecium*. *Antimicrob Agents Chemother* 2019; **63:** 3–5.

19 Kessel J, Bender J, Werner G, et al. Risk factors and outcomes associated with the carriage of tigecycline- and vancomycin-resistant *Enterococcus faecium*. *J Infect* 2021; **82:** 227–34.

20 Kwak S, Choi J, Hink T, et al. Impact of investigational microbiota therapeutic RBX2660 on the gut microbiome and resistome revealed by a placebo-controlled clinical trial. *Microbiome* 2020; **8:** 125.

21 Zaheer R, Cook SR, Barbieri R, et al. Surveillance of *Enterococcus* spp. reveals distinct species and antimicrobial resistance diversity across a One-Health continuum. *Sci Rep* 2020; **10:** 3937.

22 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 Genom* 2016; **2:** 1–11.

23 Rushton-Green R, Darnell RL, Taiaroa G, Carter GP, Cook GM, Morgan XC. Agricultural origins of a highly persistent lineage of vancomycin-resistant *Enterococcus faecalis* in New Zealand. *Appl Environ Microbiol* 2019; **85:** 1–17.

24 Wang G, Yu F, Lin H, et al. Evolution and mutations predisposing to daptomycin resistance in vancomycin-resistant *Enterococcus faecium* ST736 strains. *PLoS One* 2018; **13:** e0209785.

25 Wardenburg KE, Potter RF, D'Souza AW, et al. Phenotypic and genotypic characterization of linezolid-resistant *Enterococcus faecium* from the USA and Pakistan. *J Antimicrob Chemother* 2019; **74:** 3445–52.

26 Kerschner H, Cabal A, Hartl R, et al. Hospital outbreak caused by linezolid resistant *Enterococcus faecium* in Upper Austria. *Antimicrob Resist Infect Control* 2019; **8:** 150.

27 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:** 1–15.

28 Gouliouris T, Coll F, Ludden C, et al. Quantifying acquisition and transmission of *Enterococcus faecium* using genomic surveillance. *Nat Microbiol* 2021; **6:** 103–11.

29 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.

30 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.

31 Gouliouris T, Raven KE, Moradigaravand D, et al. Detection of vancomycin-resistant *Enterococcus faecium* hospital-adapted lineages in municipal wastewater treatment plants indicates widespread distribution and release into the environment. *Genome Res* 2019; **29:** 626–34.

32 Roberts LW, Hoi LT, Khokhar FA, et al. Genomic characterisation of multidrug-resistant *Escherichia coli, Klebsiella pneumoniae,* and *Acinetobacter baumannii* in two intensive care units in Hanoi, Viet Nam: a prospective observational cohort study. *Lancet Microbe* 2022; **3:** e857–66.

33 Alcock BP, Raphenya AR, Lau TTY, et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res* 2020; **48:** D517–25.

34 Gagnon S, Lévesque S, Lefebvre B, Bourgault AM, Labbé AC, Roger M. *vanA*-containing *Enterococcus faecium* susceptible to vancomycin and teicoplanin because of major nucleotide deletions in Tn1546. *J Antimicrob Chemother* 2011; **66:** 2758–62.

35 Lauderdale T-L, McDonald LC, Shiau YR, et al. Vancomycin-resistant enterococci from humans and retail chickens in Taiwan with unique VanB phenotype-*vanA* genotype incongruence. *Antimicrob Agents Chemother* 2002; **46:** 525–27.

36 Pietta E, Montealegre MC, Roh JH, Cocconcelli PS, Murray BE. *Enterococcus faecium* PBP5-S/R, the missing link between PBP5-S and PBP5-R. *Antimicrob Agents Chemother* 2014; **58:** 6978–81.

37 Wagner TM, Howden BP, Sundsfjord A, Hegstad K. Transiently silent acquired antimicrobial resistance: an emerging challenge in susceptibility testing. *J Antimicrob Chemother* 2023; **78:** 586–98.