Title: Improved accuracy of antibiotic resistance determination from *Enterococcus faecium* whole-genome sequences

Authors: Francesc Coll[1], Theodore Gouliouris[2,3], Beth Blane[3], Kathy E. Raven[3], Catherine Ludden[4], Fahad A. Khokhar[5], Hayley J. Wilson[5], Leah W. Roberts[3,6], Ewan M. Harrison[3,7], Carolyne S. Horner[8], Le Thi Hoi[9,10], Nguyen Thi Hoa[9,11], Nguyen Vu Trung[9,10], Nguyen Van Kinh[9], Nicholas M. Brown[2,8,12], Mark A. Holmes[5], Julian Parkhill[5], M. Estee Török[2], Sharon J. Peacock[2,3]

Affiliations:

[1] Department of Infection Biology, Faculty of Infectious & Tropical Diseases, London School of Hygiene & Tropical Medicine, London, UK

[2] Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK

[3] Department of Medicine, University of Cambridge, Cambridge, UK

[4] Genomics England, London, UK

[5] Department of Veterinary Medicine, University of Cambridge, Cambridge, UK

[6] European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Hinxton, UK

[7] Wellcome Sanger Institute, Hinxton, Cambridge, UK.

[8] British Society for Antimicrobial Chemotherapy, Birmingham, UK

[9] National Hospital for Tropical Diseases, Hanoi, Viet Nam

[10] Hanoi Medical University, Hanoi, Viet Nam

[11] Department of Microbiology and National Tuberculosis Reference Laboratory, National Lung Hospital, Hanoi, Viet Nam

[12] UK Health Security Agency, Cambridge, UK.

Corresponding authors: Francesc Coll. Email: Francesc.Coll@lshtm.ac.uk, Theodore Gouliouris (theo.gouliouris@doctors.org.uk)

Keywords: enterococci; *Enterococcus faecium*; antimicrobial resistance; antibiotic resistance; antibiotic susceptibility testing; whole-genome sequencing

ABSTRACT

Background: DNA sequencing could become an alternative to *in vitro* antibiotic susceptibility testing (AST) methods for the detection of bacterial antibiotic resistance. This is exemplified by the success in predicting antibiotic resistance (ABR) from the genomes of bacteria such as *Mycobacterium tuberculosis*, *Staphylococcus aureus* or *Streptococcus pneumoniae*. Here, we explored the feasibility of determining antibiotic susceptibility with high accuracy from *Enterococcus faecium* genomes.

Methods: We conducted a literature search to compile a catalogue of genes and mutations predictive of antibiotic resistance in *E. faecium*. We evaluated the diagnostic performance of this database to determine susceptibility to 15 different antibiotics using a dataset of 4,730 *E. faecium* isolates with available whole-genome sequences and *in vitro* culture-based AST phenotypes. We additionally used CARD, ResFinder, AMRFinder and LRE-Finder to assess their prediction performance against phenotypic resistance.

Results: Our results show that very accurate genotypic predictions can be obtained for ampicillin, ciprofloxacin, vancomycin and linezolid resistance, with sensitivity and specificity well above 95%. High sensitivity (above 90%) was also obtained for clindamycin, erythromycin, tetracyclines, teicoplanin, and aminoglycosides, although at lower specificity

(60 to 90%). Sensitivity was expectedly lower for daptomycin (26.6%) and tigecycline (38.3%), for which the mechanisms of resistance are less well understood. CARD, AMR-Finder and ResFinder produced less accurate predictions than our curated database for most antibiotics, and particularly in terms of specificity.

Conclusion: This work represents the largest published effort to date at evaluating the accuracy of antibiotic susceptibility predictions from *E. faecium* genomes. Our findings demonstrated an improved diagnostic accuracy compared to available ABR databases and bioinformatic tools, and provide new information for future improvements on ABR genotype–phenotype concordance in *E. faecium*.

RESEARCH IN CONTEXT

Evidence before this study

Early proof-of-concept studies demonstrated the feasibility of determining antibiotic resistance from bacterial genome sequences. Population-based studies are needed to assess the diagnostic accuracy of genotypic predictions compared to conventional phenotypic AST methods. We conducted a literature search to identify studies employing whole-genome sequencing for the detection of antibiotic resistance 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 December 2022. Out of the 161 articles retrieved, 5 investigated the concordance between WGS-based predictions and phenotypic AST results. One study investigated the prediction of linezolid resistance specifically. The rest considered predictions for 7 to 13 antibiotics, using relatively modest sample sizes (100 to 205 *E. faecium* isolates), low numbers (<15) of resistant isolates for some antibiotics (e.g. tigecycline), or low numbers of susceptible isolates for others (e.g. ampicillin). These studies used ResFinder, LRE-finder, or custom catalogues of resistance markers, generally followed by expert curation and interpretation of the results, to establish concordance between genotypic and phenotypic AST results.

Added value of this study

In this study we curated a catalogue of 297 genetic markers involved in resistance to 15 different antibiotics in *E. faecium*. We used a large and diverse collection of 4,730 *E. faecium* isolates, with available whole-genome sequences and conventional AST phenotypes, to assess the diagnostic accuracy of genotypic predictions obtained by our curated database. Nearly complete genotype-phenotype concordance could be achieved for ampicillin, ciprofloxacin,

vancomycin and linezolid. High sensitivity was also obtained for the detection of resistance to clindamycin, erythromycin, tetracyclines, teicoplanin, and aminoglycosides, although with lower specificity. Sensitivity was lower for the detection of daptomycin and tigecycline resistance. We additionally assessed the genotypic predictions of CARD, ResFinder, and AMRFinder, and found an overall improved accuracy of our curated database compared to these tools and databases. We have made our curated database of AMR genetic markers, collection of whole-genome sequences and AST data, and underlying code and bioinformatic pipelines publicly available to ensure reusability of our results.

Implications of all the available evidence

Our study represents an important improvement in the accuracy of antibiotic resistance determination from *E. faecium* genomes. Given the mortality burden of antibiotic-resistant *E. faecium* including vancomycin-resistant *E. faecium* (VRE) and the increasingly routine use of whole-genome sequencing in clinical microbiology labs, the results and resources (dataset and curated AMR database) made available here will facilitate the adoption of whole-genome sequencing as a tool for the diagnosis and surveillance of AMR in *E. faecium*.

INTRODUCTION

The World Health Organization (WHO) has declared antimicrobial resistance (AMR) as one of the top 10 global public health threats facing humanity. Without effective antimicrobials, the success of modern medicine in treating infections, including during major surgery and cancer chemotherapy, would be at increased risk. Antibiotic-resistant *Enterococcus faecium*, especially vancomycin-resistant *E. faecium* (VRE), are of particular concern. *E. faecium* is a leading cause of hospital-acquired infections such as bloodstream, surgical site, or urinary tract infections, especially in immunocompromised and critically ill patients. Healthcare-associated *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 mortality burden, associated with an estimated 200,000 deaths per annum globally.[1] In Europe, the proportion of vancomycin-resistant *E. faecium* (VRE), among all invasive *E. faecium* isolates, increased from 8.1% in 2012 to 19.0% in 2018.[2] As such, VRE is regarded as a high priority pathogen by the WHO.[3]

Growing rates of AMR make antibiotic susceptibility testing (AST) increasingly needed to ensure that the right antibiotics are prescribed to treat bacterial infections. AST is routinely performed using culture-based methods such as disk diffusion, broth microdilution or gradient diffusion (the E-test). As antibiotic resistance is genetically encoded, molecular tests have been developed to detect genetic markers of AMR. Whole-genome sequencing (WGS) represents an attractive alternative to both culture and targeted molecular tests for the detection of AMR as it can, in principle, detect the entire repertoire of AMR genetic determinants (genes and mutations) and predict resistance to virtually all antibiotics in a single experiment.[4–6] The accuracy of such predictions depends on the availability of curated and updated databases of AMR genetic determinants, and large and diverse collections of isolates with available 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.[7–11]

The success in predicting phenotypic resistance from whole-genome sequences is well exemplified for several other bacterial species. For *Mycobacterium tuberculosis*, genotypic predictions of susceptibility to first-line drugs correlate well with phenotypic susceptibility.[10,12,13] In *Staphylococcus aureus*, previous studies have shown nearly perfect concordance between phenotypic AST results and genotypic resistance predictions,[10,14] as is also the case for *Streptococcus pneumoniae*.[15] For *E. faecium*, previous studies[4,9,16–19] used isolate collections of small sample sizes (100 to 205 isolates) to assess genotype–phenotype concordance for seven to thirteen antibiotics, and with relatively low resistance rates for some antibiotics.

In this work, we conducted a literature review to assemble a catalogue of acquired genes and mutations predictive of antibiotic resistance in *E. faecium*. We evaluated the diagnostic performance of this database to predict ABR to 15 different antibiotics using a dataset of 4,730 *E. faecium* isolates, with both culture-based antibiotic susceptibility phenotypes and whole-genome sequences available. Our results show very accurate susceptibility predictions for most antibiotics, an improved accuracy compared to other available ABR databases and bioinformatic tools and clear directions to achieve full ABR genotype-phenotype concordance.

RESULTS

Collections of whole-genome sequenced E. faecium isolates used in this study

We assembled a dataset of 4,730 *E. faecium* isolates from 24 independent studies[4,9,27–36,18,37,38,20–26] with available whole-genome sequences and antibiotic susceptibility phenotypes. Table 1 presents the number of isolates, antibiotics tested, AST method used and source of isolates per study, and Supplementary Data 1 provides isolate-level metadata. The phenotypic AST data of 3,435 (72.6%) of these *E. faecium* isolates, whose genome sequences had been already published,[34–38] were made available for this study. Overall, we obtained susceptibility phenotypes for 15 different antibiotics: 340 isolates tested for daptomycin (37.4% non-susceptible), 359 for doxycycline (38.4% non-susceptible), 673 for ciprofloxacin (74.4% non-susceptible), 2,154 for clindamycin (71.1% non-susceptible), 2,338 for quinupristin/dalfopristin (33.7% non-susceptible), 2,364 for tetracycline (79.3% non-susceptible), 2,412 for tigecycline (2.5% non-susceptible), 2,635 for streptomycin (43.7% non-susceptible), 2,640 for kanamycin (72.9% non-susceptible), 2,760 for teicoplanin (71.8% non-susceptible), 3,050 for linezolid (4.8% non-susceptible), 3,125 for gentamicin (51.5% non-susceptible), 3,176 for ampicillin (89.3% non-susceptible), 3,251 for erythromycin (97.1% non-susceptible), and 3,409 for vancomycin (68.1% non-susceptible). See Methods for definition of non-susceptibility phenotypic category, and Supplementary Tables 1 and 2 for MIC and zone diameter breakpoints used.

Database of genetic determinants of antibiotic resistance in E. faecium

We compiled a curated database of 297 genetic determinants comprising single mutations (n=103), combinations of mutations (n=100), chromosomal genes (n=4), single acquired genes (n=64), and multiple acquired genes (i.e. operons, n=26) for a total of 17 antibiotics, that is, the 15 antibiotics selected plus streptogramin A and streptogramin B. We encoded for cross-

resistance across tetracyclines (i.e. doxycycline and tetracycline). Resistance to quinupristin/dalfopristin was called if isolates contained resistance determinants for both streptogramin A and streptogramin B. As individual genetic determinants (of any of the five categories above) can influence susceptibility to one or multiple antibiotics, susceptibility to each of them was encoded as a separate relationship and defined by an 'effect' on susceptibility which we classified as 'resistance' (n=212), 'susceptibility' (n=4), or 'undetermined' (n=81), based on the existence or lack of evidence in the scientific literature (see Methods). Only 'resistance' and 'susceptibility' categories were used to make genotypic predictions. Table 2 summarizes the genetic determinants with 'resistance' effect included in the curated database. Supplementary Data 2 includes the complete database, including the genetic determinants with 'undetermined' effect.

Genotypic predictions of antibiotic resistance using the curated database

Table 3 summarizes the diagnostic performance in terms of sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV), of using the curated database to determine phenotypic resistance. We calculated these diagnostic metrics to predict both phenotypic resistance (Table 3) and phenotype non-susceptibility (Supplementary Table 3) (See Methods for definition of these phenotypic categories).

Ampicillin resistance (> 8 µg/mL MIC) was predicted with very high sensitivity (99.7%) and specificity (97.9%). Ampicillin non-susceptibility (> 4 µg/mL MIC) was predicted with the same specificity (97.9%) and very high sensitivity too (99.6%, Supplementary Table 3), due to the presence of ampicillin-resistant *pbp5* mutations in isolates with MIC of 8 µg/mL (Figure 1A). Ampicillin resistance was attributable to *pbp5* mutations only, no isolate carried the *blaZ* operon. The *pbp5* mutation E629V, proven to cause mild increases in ampicillin MIC [39] was

the most frequently detected mutation in non-susceptible isolates (n=2,825/2,836, 99.6%), usually found in combination with M485A (n=2,693, 95.0%) or M485T (n=68, 2.4%). Mutations V586L (n=391, 13.8%) and M426I (n=1,415, 50.0%) were also found in combination with these three mutations in non-susceptible isolates (Supplementary Figure 1). *pbp5* mutations E525D, N496K, A499T and A401S, previously suspected to be involved in ampicillin resistance, were often found in susceptible isolates and consequently discarded as genetic determinants of resistance.

Ciprofloxacin resistance (> 4 µg/mL MIC) was predicted with very high sensitivity (98.0%) and specificity (98.8%), and attributable to combinations of *gyrA* and *parC* mutations, the most common being combinations of *gyrA* S84 and *parC* S82 amino acid changes (n=465/501, 92.8%) and less often the combination of *gyrA* E88G and *parC* S82I (n=26, 5.2%).

Vancomycin resistance (> 4 µg/mL MIC) was predicted with very high sensitivity (98.3%) and specificity (98.8%). The *vanA* or *vanB* operons were detected in 2,187 (94.3%) and 88 (3.8%) isolate, and both operons in six isolates (0.2%), out of the 2,320 vancomycin-resistant isolates. Because *vanA*-positive vancomycin-susceptible isolates have been described,[40] caused by the loss of certain genes in the *vanA* operon, we encoded resistance based on 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 made up of all seven genes (*vanA*, $vanH_A$, $vanR_A$, $vanS_A$, $vanX_A$, $vanY_A$, $vanZ_A$) but also for variants of the *vanA* operon lacking certain genes (Table 2), as we found such variants in phenotypically resistant isolates (Figure 2). The full *vanA* operon was recovered in most resistant cases (1,911, 82.4%); followed by vanAΔ$vanY_A$, i.e. the vanA operon variant lacking $vanY_A$ (175, 7.5%); the full *vanB* operon (94, 4.1%, which include six

isolates carrying the *vanA* operon too); vanAΔ*vanY*$_A$Δ*vanZ*$_A$ (58, 2.6%); vanAΔ*vanZ*$_A$ (36, 1.4%); vanBΔ*vanY*$_B$ (8, 0.3%); and vanAΔ*vanS*$_A$ (6, 0.3%). If vancomycin resistance had been called based on the detection of *vanA* or *vanB* genes alone, this would have led to a 5.5% drop in specificity (93.3%), and just a marginal 0.3% increase in sensitivity (98.6%).

Teicoplanin resistance (> 2 µg/mL MIC) was predicted with very high sensitivity (98.9%) but lower specificity (93.3%). The full *vanA* operon was detected in most resistant cases (1,708/1,982, 86.2%), followed by vanAΔ*vanY*$_A$ (174, 8.8%); vanAΔ*vanY*$_A$Δ*vanZ*$_A$ (47, 2.4%); vanAΔ*vanZ*$_A$ (29, 1.4%); and vanAΔ*vanS*$_A$ (3, 0.1%). Specificity was low due to 52 isolates carrying a *vanA* operon variant but being phenotypically susceptible to teicoplanin. The majority of these false positives (44/52) were phenotypically resistant to vancomycin (> 4 µg/mL MIC), thus displaying what has been previously described as a VanB phenotype-*vanA* genotype.[41–43] Of these 44 isolates, 12 isolates from the same collection[35] were re-tested using E-test (Supplementary Data 3) and found to exhibit an heteroresistance phenotype, as showed by colonies within the resistant zone on the E-test strips.

Erythromycin resistance (> 4 µg/mL MIC) was predicted with very high sensitivity (94.3%) and specificity (96.8%). Phenotypic resistance (n=2,824) could be explained by the acquisition of single and combinations of *erm* genes, including *ermB* alone (n=1,623); *ermT* alone (n=383); *ermB* plus *ermT* (n=288); *ermB* plus *ermA* (n=188); *ermB*, *ermT* plus *ermA* (n=98); *ermA* alone (n=72); and *ermT* plus *ermA* (n=18). However, we found an unexpectedly high number of false negatives for erythromycin (n=160, Table 3), so we re-tested five randomly selected false negatives clustered in large phylogenetic clades (presumed more likely to carry a novel resistance mechanism) and five sporadic false negatives (presumed more likely to represent a

phenotypic testing error). See Methods for re-testing selection criteria. All five sporadic false negatives tested susceptible to erythromycin with E-test (Supplementary Data 3) indicating an incorrect original Vitek result, while two out of the five clustered false negatives displayed a variable susceptible/resistant phenotype, pointing to a potential novel erythromycin resistance mechanism in a subset of clustered false negatives.

Resistance to doxycycline and tetracycline was predicted with very high sensitivity (97.0% and 99.5%, respectively) but at low specificity, 81.0% and 71.0% respectively (Table 3). Tetracycline resistance (n=1,869) was caused by single and combinations of acquired *tet* genes, being the combination of *tetL* plus *tetM* the most common (n=1,049) followed by *tetM* alone (n=439), *tetM* plus *tetS* (n=198), *tetL* alone (n=73), *tetS* alone (n=72), the combination of *tetL*, *tetM* plus *tetS* (n=26); the combination of *tetL*, *tetM* plus *tetO* (n=2) and *tetO* alone (n=1). Doxycycline resistance (n=99) was caused the combination of *tetL* plus *tetM* (n=76) and *tetM* alone (n=20). Previous studies have also reported very high false positive rates for tetracyclines in *E. faecium*,[16,17] caused by the detection of *tet* genes in phenotypically susceptible isolates. The low specificity observed could be attributable to isolates with an incorrect pAST result or the presence of silenced AMR genes. To clarify this, we plotted the distribution of false positives on the phylogenetic tree (Supplementary Figure 2) and re-tested a representative subset of these. As the high degree of clustering of false positives may be indicative of silenced *tet* genes, and a genetic basis for this, we re-tested five false positives clustered in large clades of false positives (presumably more likely to represent cases of AMR gene silencing), 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). As hypothesised, all five clustered false positives displayed a susceptible phenotype using E-test (Supplementary Data 3), indicating that their *tet* genes are silenced, while four of the five sporadic false

299 positives tested resistant on E-test, indicating an erroneous original Vitek result. Altogether,
300 these results point to phenotypic testing errors and silenced *tet* genes as major sources of false
301 positive genotypic predictions for tetracyclines.

302

303 High-level aminoglycoside resistance was predicted with very high sensitivity, with 97.7%,
304 96.8% and 99.4% for streptomycin, gentamicin and kanamycin, respectively (Table 3). False
305 positive rates were high, due to the presence of aminoglycoside-resistance genes in isolates
306 with MICs below the high-level resistance breakpoint (i.e. > 512, > 128 and > 512 μg/mL
307 MICs, respectively), resulting in genotypic predictions of low specificity (88.6%, 82.2% and
308 66.1%, respectively). High-level streptomycin resistance (n=1,152) was conferred by the
309 acquisition of both *aad(6)* and *ANT(6)-Ia* genes in most isolates (n=1,085), followed by
310 *ANT(6)-Ia* alone (n=37) or *aad(6)* alone (n=3); high-level gentamicin resistance (n=1,610) by
311 the acquisition of *AAC(6')-Ie-APH(2'')-Ia* (n=1,550) or *APH(2'')-IVa* (n=8); and high-level
312 kanamycin resistance (n=1,924) by the acquisition of both *AAC(6')-Ie-APH(2'')-Ia* and
313 *APH(3')-IIIa* (n=830), *APH(3')-IIIa* alone (n=614), *AAC(6')-Ie-APH(2'')-Ia* alone (n=462),
314 *APH(2'')-IVa* plus *APH(3')-IIIa* (n=5) or *APH(2'')-Iva* (n=2). The low specificity observed
315 could be attributable to isolates with an elevated MIC below the high-level resistance
316 breakpoint, a wrong pAST result or the presence of silenced AMR genes. To investigate this,
317 we focused on isolates with available gentamicin MICs (n=534). A limitation of our dataset is
318 that aminoglycoside MICs were not available for most isolates. Of the 69 false positives with
319 available gentamicin MIC, 33 had MICs above the ECOFF (>32 μg/mL) and 36 below
320 (Supplementary Figure 3), the latter carrying putatively silenced AMR genes. To confirm the
321 presence of silenced AMR genes, we re-tested 10 false positives with Vitek MICs below the
322 ECOFF (≤32 μg/mL) using E-test, including five clustered and five sporadic cases. The E-test
323 results (Supplementary Data 3) showed a mixture of gentamicin resistant (>32 μg/mL, n=7)

and susceptible (≤32 μg/mL, n=3) isolates, the latter indicating the presence of silenced aminoglycoside-resistance genes.

Clindamycin resistance was predicted with high sensitivity (97.1%) but low specificity (62.6%) (Table 3). Resistance (>2 μg/mL, n=1531) was attributable to combinations of acquired *erm* genes, *lseE* and *eatA* T450I mutation, including *eatA* T450I plus *ermB* (n=535/1,531, 34.9%), *ermB* alone (432, 28.2%), *ermB* and *ermA* (157, 10.3%), *eatA* T450I alone (121, 7.9%), *ermB* plus *ermA* plus *eatA* T450I (86, 5.6%), *ermB* plus *lsaE* plus *eatA* T450I (68, 4.4%), and other combinations of these four determinants (87, 5.7%). False positives were caused by the detection of *ermB* (n=110, 17.7%), *ermA* (69, 11.1%), *eatA* T450I (42, 6.7%) and *ermB* plus *ermA* (9, 1.4%) in phenotypically susceptible isolates (n=623).

Resistance to the antibiotic combination quinupristin-dalfopristin was called if isolates carried genetic determinants of both streptogramin A (dalfopristin) and streptogramin B (quinupristin) (Table 2). The sensitivity and specificity of genotypic predictions were high (89.2% and 94.5%, respectively), although lower than that achieved for other antibiotics (Table 3). Phenotypic resistance (> 4 μg/mL, n=314) was explained by the coincident detection of *eatA* T450I mutation, conferring resistance to dalfopristin, and *ermB*, conferring resistance to quinupristin (n=280, 89.2%). Eight of these 280 isolates also carried *vatE*. Phenotypic non-susceptibility (> 1 μg/mL, n=788) was predicted with lower sensitivity (81.1%) and the same specificity (Supplementary Table 3).

The sensitivity of genotypic resistance predictions to the last-line antibiotics tigecycline, daptomycin and linezolid were the lowest (38.3%, 72.4% and 79.6%) compared to all other

antibiotics investigated (Table 3), albeit specificities were high (99.7%, 92.0% and 98.2%). Resistance to tigecycline (n=60) was conferred by non-synonymous SNPs in *rpsJ*, specifically by K57E, K57R and Y58D in 13, six and four resistant isolates, respectively. Daptomycin resistance was conferred by diverse combinations of mutations in *liaR*, *liaS* and *cls* genes, including *liaS* (n=13), *liaR* (n=4) or *cls* mutations alone (n=5); and combinations of *liaS* plus *liaR* (n=49), *liaS*, *liaR* plus *cls* (n=18), or *liaS* plus *cls* mutations (n=3). Overall, the daptomycin-conferring mutations detected were *liaS* T120A (n=81), *liaR* W73C (n=69), *cls* R218Q (n=9), *cls* H215R (n=8), *cls* A20D (n=7) and *cls* N13I (n=2). Linezolid resistance (n=147) was conferred by the G2589T SNP in 23S rRNA (n=72), the acquisition of both *optrA* and *poxtA* genes (n=34), *poxtA* alone (n=5), *cfrB* alone (n=5) or G2518A SNP in 23S rRNA (n=1). We detected an unexpectedly high number of false negatives for linezolid (n=30) whose phenotypic resistance should have been explained by known resistance mechanisms. We therefore re-tested all false negatives (Supplementary Data 3) using E-test, which displayed a susceptible linezolid phenotype (≤4 μg/mL). When considering these new phenotypic results, the sensitivity of genotypic predictions for linezolid became 100% (Table 3).

In summary, resistance to ampicillin, ciprofloxacin, vancomycin and linezolid was determined with very high sensitivity and specificity (at or above 98%). Very high sensitivity was achieved for teicoplanin, clindamycin, tetracyclines and high-level aminoglycoside resistance, albeit at a lower specificity. Sensitivity still needs improvement for erythromycin, quinupristin-dalfopristin and, as expected, for the last-line antibiotics tigecycline and daptomycin, for which their mechanisms of resistance are not fully understood.

Effect of different breakpoints on genotypic predictions

We compared the effect of using either CLSI or EUCAST clinical breakpoints (Supplementary Table 1) on the diagnostic accuracy of genotypic predictions (Table 3 and Supplementary Table 4). We limited our comparison to the antibiotics with breakpoints set by both organizations as CLSI breakpoints are not set for tigecycline, kanamycin and clindamycin; and EUCAST clinical breakpoints are not available for daptomycin, tetracycline, doxycycline, kanamycin, erythromycin and clindamycin.

Ampicillin, ciprofloxacin and vancomycin resistance were predicted with very high sensitivity (99.7%, 98.0%, 98.3%) and specificity (97.9%, 98.8%, 98.8%) when using EUCAST's clinical breakpoints (> 8, > 4 and > 4 μg/mL). For ampicillin, the use of CLSI breakpoints (Supplementary Table 4), which classify isolates of 8 μg/mL MIC as susceptible, led to a decrease in specificity (91.5%) due to the presence of ampicillin-resistant *pbp5* mutations in isolates with MIC of 8 μg/mL (Figure 1A). For ciprofloxacin, CLSI breakpoints led to a reduction in sensitivity (88.0%) due to isolates with 4 μg/mL MIC being classified as phenotypically resistant but lacking any ciprofloxacin resistance mutations. The use of CLSI breakpoints yielded very similar performance for vancomycin, but worse specificity (84.8%) for teicoplanin (> 16 μg/mL MIC) than EUCAST breakpoints (> 2 μg/mL MIC, 93.3% specificity). For quinupristin-dalfopristin, CLSI breakpoints and EUCAST produced very similar specificity (94.4% vs. 94.5%) and sensitivity (86.9% vs. 89.2%). CLSI and EUCAST breakpoints are the same for linezolid, and high-level streptomycin and gentamicin resistance.

In summary, the use of EUCAST clinical breakpoints resulted in better phenotype-genotype concordance for ampicillin, ciprofloxacin and teicoplanin resistance.

Diagnostic performance of genotypic predictions obtained by other publicly available tools

As a comparison, we also determined antibiotic resistance from *E. faecium* genomes using the latest versions of AMRFinderPlus[7] (Supplementary Table 5), ResFinder[8] (Supplementary Table 6), RGI CARD[44] (Supplementary Table 7) and LRE-Finder[9] for linezolid resistance (Supplementary Table 8). Relationships of cross-resistance were assumed among antibiotics of the same antibiotic class when individual antibiotics were not explicitly predicted by an AMR database (Supplementary Table 9). These tools produced comparable sensitivities and specificities (Figure 3) for ciprofloxacin (ResFinder and AMRFinderPlus), linezolid (LRE-Finder), and high-level aminoglycoside resistance (AMRFinderPlus), although they most often under-performed compared to our curated database. 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* (vanAΔ*vanR*$_A\Delta$*vanS*$_A$) and *vanB* (vanBΔ*vanR*$_B\Delta$*vanS*$_B$) operons detected in phenotypically susceptible isolates. For tetracycline, the lower specificity achieved by ResFinder and AMRFinderPlus is explained by the detection of a truncated variant of *tetM* (a frameshift mutation at K261fs) in phenotypically susceptible isolates.

Very low specificity was 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. Similarly, the specificity of clindamycin resistance predictions by ResFinder and CARD was low due to the detection of *ermT* in phenotypically susceptible isolates, a gene known to cause resistance to macrolides but not to lincosamides in *E. faecium*. Analogously, the inclusion of *msrC* as a macrolide-resistance determinant in ResFinder and CARD led to a drop in specificity when predicting erythromycin resistance.

msrC is a chromosomally-encoded gene intrinsic to *E. faecium* strains[45] which is unlikely, on its own, to confer MIC increases above the resistance breakpoint (> 4 μg/mL). CARD produced very low specificity for most antibiotics (Figure 3).

Very high specificity in predictions, comparable 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 sensitivities were lower than those obtained by the curated database. Lower sensitivity was caused by the absence of known resistance-conferring determinants in these databases present in phenotypically resistant isolates. The diagnostic performance of LRE-Finder to predict linezolid resistance (Supplementary Table 8) was comparable to that obtained by the curated database (Table 3).

DISCUSSION

This work represents the largest effort conducted to date at evaluating the accuracy of antibiotic resistance predictions from *E. faecium* genomes. We curated a database of 297 genetic determinants known to cause resistance to 15 different antibiotics in *E. faecium*, and evaluated the diagnostic performance of this database to predict phenotypic resistance in a dataset of 4,730 *E. faecium* genomes. We showed that very accurate resistance predictions could be made for ampicillin, ciprofloxacin, vancomycin, linezolid, teicoplanin, erythromycin and quinupristin-dalfopristin. The specificity of genotypic predictions needs to be improved for tetracyclines and high-level aminoglycoside resistance; and sensitivity 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 limited.[4,17] This was due to limited availability of *pbp5* DNA sequences, and the narrow comparison of the *pbp5*-resistance allele carried by clade A1 isolates against the *pbp5*-susceptible allele carried by clade B isolates, which differ by 21 amino acids.[46] Here, the availability of diverse clade A2 isolates with different ampicillin susceptibility (Supplementary Figure 1) allowed us to differentiate the key amino acid changes contributing to ampicillin resistance, improving our ability to predict phenotypic resistance with very high accuracy.

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 in the curated database. Calling glycopeptide resistance based on the detection of the *vanA* or *vanB* genes alone would have resulted in significant drops in specificity and marginal increases in sensitivity, due to the presence of truncated variants (e.g. vanAΔ$vanR_A$Δ$vanS_A$, vanBΔ$vanR_B$Δ$vanS_B$) of these operons in phenotypically susceptible isolates. On the other hand, calling resistance based on the detection of the full *vanA* and *vanB* operons would have resulted in lower sensitivity, as we found operon variants lacking certain genes (i.e. vanAΔ$vanY_A$, vanAΔ$vanZ_A$, vanAΔ$vanY_A$Δ$vanZ_A$ and vanAΔ$vanS_A$) in phenotypically resistant isolates. These observations highlight the importance of accounting for all the different *vanA* or *vanB* operon variants present in phenotypically resistant strains (i.e. functional), and avoid calling glycopeptide resistance based on the detection of *vanA* or *vanB* genes alone.

Commonly used AMR databases produced worse genotypic predictions than our curated database, with the exception of ciprofloxacin (ResFinder and AMRFinderPlus), linezolid

(LRE-Finder) and high-level aminoglycoside resistance (AMRFinderPlus) for which their diagnostic performance was comparable. The detection of truncated and non-functional operon variants (e.g. *vanA* and *vanB* operons), truncated variants of genes (e.g. frameshift mutation in *tetM*), non-conferring resistance mutations (e.g. in *pbp5*) and genes (*msrC*), and incorrectly encoded gene-resistance relationships (e.g. *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 (i.e. fully assembled without frameshifts or truncations) when calling resistance; and the importance of continuous and improved curation of existing databases.

Both our curated database and other tools achieved a low specificity in the prediction of tetracycline and high-level aminoglycoside resistance. In addition to erroneous phenotypic AST results being the source of part of these discrepancies, we showed that clustering of false positives in the phylogeny, i.e. the detection of clonally-related and phenotypically-susceptible isolates carrying the same resistance gene(s), was common and indicated the presence of silenced AMR genes after phenotypic re-testing. This phenomenon could represent what has previously been termed as *Silencing of Antibiotic Resistance by Mutation* (SARM) and investigated in *Staphylococcus aureus*.[47] 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 present in phenotypically resistant strains explained the lower sensitivity obtained for certain antibiotics, particularly for last-line antibiotics (i.e. tigecycline and daptomycin). Although the curated database achieved better sensitivities for these antibiotics, further research is needed to obtain

a complete understanding of the genetic basis of resistance to last-line antibiotics in *E. faecium*. This is clinically important as these last resort antibiotics are used for the treatment of VRE.

We additionally compared the effect of using either CLSI or EUCAST clinical breakpoints on the accuracy of genotypic predictions, and found that EUCAST breakpoints resulted in better phenotype-genotype concordance for ampicillin, ciprofloxacin and teicoplanin resistance. These results demonstrate the utility of genotypic resistance determinations in differentiating wildtype from non-susceptible populations, and establishing the critical concentrations to define phenotypic resistance.[48,49]

Our study has several limitations. We limited our approach to make binary predictions of antibiotic susceptibility, that is, to differentiate resistant (or non-susceptible) from susceptible strains. 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 (i.e. MICs) from genomic data, particularly to discriminate between the intermediate and resistance categories, and inform likelihood of therapeutic success. Despite the large dataset used, the number of tested isolates was modest (~200 to 400 isolates) for some antibiotics (daptomycin and doxycycline) and the resistance rates low (2.5 and 5.6%) for others (tigecycline and linezolid). Larger and more resistant *E. faecium* populations are needed to obtain more accurate performance estimates for these antibiotics.

In conclusion, we curated a database of genetic determinants of antibiotic resistance in *E. faecium* and assessed its diagnostic performance in a large and diverse collection of sequenced strains. The accuracy in predictions was equivalent or enhanced compared to existing databases. A complete characterization of the genetic basis of non-susceptibility to last-line

antibiotics, and the mechanisms mediating antibiotic resistance silencing, are needed to close the remaining sensitivity and specificity gaps in genotypic predictions.

METHODS

Literature review on genetic determinants of antibiotic resistance

We aimed at establishing the role of individual genes and mutations on the susceptibility to specific antibiotics, not to antibiotic classes. We started by extracting a list antibiotics and AMR genes from review articles on mechanisms of antibiotic resistance in enterococci[50–56] and AMR genotype-phenotype studies in enterococci.[4,9,17] We then conducted a targeted PubMed search for each antibiotic-gene combination to identify the earliest and cumulative evidence available to date on the role of each chromosomal and acquired gene in the susceptibility to individual antibiotics. For cases of mutational resistance, we search for all reported mutations, and combination of mutations, in chromosomal AMR genes.

Database on genetic determinants of antibiotic resistance in enterococci

Following the literature search, each ABR genetic determinant was classified into one of the following five categories: chromosomal gene, acquired gene, acquired multiple genes, single mutation, or mutation combination. The effect of genetic determinants on antibiotic susceptibility was classified as either 'resistance', 'susceptibility' or 'undetermined'. The effect of individual ABR determinants was classified as 'undetermined' if they: (1) were reported in enterococci species other than *E. faecium*; (2) co-occurred with other determinants in the same resistant strains so the role of individual genetic determinants could not be established; (3) were often detected in phenotypically susceptible isolates in our dataset; or (4) if mutations could not be annotated onto the reference sequence (see below). Evidence on the effect of genetic determinants was searched specifically for *E. faecium*, that is, we did not

assume determinants to have the same effect across different enterococcal species. We deliberately avoided encoding relationships between genetic determinants and antibiotic classes, and focused on individual antibiotics instead. The reference nucleotide sequence of chromosomal genes was extracted from the reference genome of *E. faecium* Aus0004 strain (GenBank accession no. CP003351), except for that of *pbp5* which was extracted from the ampicillin-susceptible *E. faecium* Com15 strain (GenBank accession no. CP025022.1)[57]. The reference nucleotide sequence of acquired genes was downloaded preferentially from CARD v3.1.0[44], followed by ResFinder v4.0 (database version downloaded in April 2020)[58,59] if not available in CARD, and from the NCBI GenBank if not available in any of the two. See Supplementary Data 2 for full list of gene names and sequences. We transferred mutations to the Aus0004 gene numbering system by comparing the reference allele (nucleotide or amino acid) reported in original publications with that in the Aus0004 refence sequence. If they did not match, mutation sites (i.e. nucleotide or codon position) were transferred to the Aus0004 numbering system, usually annotated originally in the *E. coli* numbering system or in another enterococcal strain. In a few cases mutations could not be annotated onto the reference sequence and were labelled as 'undetermined' and not used to call resistance. Mutations were annotated following GARC rules (http://fowlerlab.org/2018/11/25/goarc-a-general-ontology-for-antimicrobial-resistance-catalogues/), with a slight adaptation for indels, to include both the reference and alternative alleles. For example, mutation pbp5@T499I indicate a single amino acid substitution at codon 499 on *pbp5*; 23SrRNA@g2589t a single nucleotide substitution at nucleotide position 2589 in gene 23S rRNA; and rpsJ@I51_del_IRATH a deletion of amino acids IRATH at amino acid positions 52-56, where the reference amino acid at codon 51 is an I.

Collections of whole-genome sequenced enterococcal strains used in this study

We used publicly available collections of *E. faecium* from published studies[4,9,27–33,18,20–26] for which whole-genome sequences (provided as either short reads or assemblies) and antibiotic susceptibility phenotypes were available (n=1,295). These studies were identified by querying the NCBI Short Read Archive (SRA) on January 2021 to identify BioProjects that contained *E. faecium* genomic sequences, could be linked to a publication, included genomes of natural isolates (i.e. of clinical, animal or environmental origin, and not laboratory-evolved or genetically-modified strains), and had isolate-level phenotypic AST metadata. The phenotypic AST data of an additional 3,435 *E. faecium* isolates, whose genome sequences had been already published[34–38,60] were made available for this study. Categorical antibiotic susceptibility labels (i.e. susceptible, intermediate or resistant), the AST method used and minimum inhibitory concentrations (MICs) or disk diameters measurements reported in these studies were all collected. We kept 15 antibiotics that had at least 50 susceptible and 50 resistant isolates available across studies: ampicillin, ciprofloxacin, clindamycin, daptomycin, doxycycline, erythromycin, gentamicin, kanamycin, streptomycin, linezolid, quinupristin-dalfopristin, tetracycline, tigecycline, teicoplanin and vancomycin.

Standardization of phenotypic antibiotic susceptibility data across studies

We extracted the MIC and zone diameter breakpoints used to define resistance and non-susceptibility in *E. faecium* for the fifteen antibiotics studied, including EUCAST epidemiological cut-offs,[61] EUCAST clinical breakpoints v12.0[62] and CLSI breakpoints.[63] Non-susceptibility is defined as the category comprising both the intermediate (or the EUCAST category 'Susceptible, increased exposure') and the resistant categories. EUCAST clinical breakpoints were chosen preferentially (Supplementary Table 1 and 2). For antibiotics with no EUCAST clinical breakpoints set (i.e. daptomycin, tetracycline, doxycycline, and

596 erythromycin), we used CLSI breakpoints instead. For kanamycin and clindamycin, whose
597 breakpoints are not set by either EUCAST or CLSI, we used the breakpoints proposed
598 elsewhere.[4] MIC values and zone diameter measurements were used to assign susceptible,
599 resistance or non-susceptible labels to isolates across studies using the selected MIC and zone
600 diameter breakpoints. See Table 1 for the antibiotics tested in each study and AST method
601 used. Supplementary Data 1 contains the original and standardised antibiotic susceptibility
602 labels.

603

604 *In silico prediction of antibiotic resistance from whole-genome sequences*

605 We used ARIBA[64] for the identification of antibiotic resistance genes and mutations from
606 whole-genome sequences as it is an established, open source bioinformatic tool designed for
607 that purpose. We thus formatted our curated database of AMR genetic determinants to ARIBA-
608 complaint input files, that is, to a metadata and FASTA file of reference sequences using a
609 custom python script (prepare_ariba_files_from_amr_database.py). ARIBA v2.14.6 was run
610 separately on each isolate genome. For isolates with only genome assemblies available, perfect
611 paired-end fastq reads were generated from assemblies using *fastaq* v3.17.0
612 (https://github.com/sanger-pathogens/Fastaq). A custom python script was written to parse
613 ARIBA reports and call resistance (call_resistance_from_ariba_report.py) based on the
614 presence of single or combinations of mutations, and on the presence and integrity (i.e. fully
615 assembled genes without frameshifts or truncations) of single or multiple acquired genes. In
616 this script, we used a minimum ratio of reads supporting alternative alleles of 10% to call
617 resistance from heterozygous mutations. The default 10% cut-off was chosen to allow the
618 detection of a single mutated 23S allele when predicting linezolid resistance, while ignoring
619 sequencing errors, as previously proposed.[9] We additionally run the command-line tools of
620 LRE-Finder[9] v1.0.0, ResFinder[8] v4.1.10 (which includes PointFinder, version downloaded on

March 2022), AMRFinderPlus[7] v3.10.18 with database version 2021-12-21.1, and RGI v5.2.1 using CARD database v3.1.4[44], to obtain genotypic predictions of antibiotic resistance for all isolate genomes. All scripts can be found in the GitHub page of this project (https://github.com/francesccoll/amr_efm_wgs).

Assessing the diagnostic performance of genotypic predictions

The selected MIC and zone diameter breakpoints (Supplementary Tables 1 and 2) were used to assign either the 'susceptible' or 'resistance' label to all tested isolates for each antibiotic. In this case, the phenotypic intermediate category was ignored. We additionally assigned the label 'susceptible' or 'non-susceptible' to all isolates, wherein the 'non-susceptible' category comprises both the intermediate and resistant categories. These phenotypic categories were used as the reference standard and compared against genotypic predictions made using our database of curated AMR genetic determinants. Additionally, phenotypic categories were also compared to genotypic predictions made by ResFinder[8] v4.1.10, AMRFinderPlus[7] v3.10.18 and RGI v5.2.1 CARD v3.1.4.[44]

We classified isolates as: true positives, if phenotypically resistant (or non-susceptible) with known resistance-conferring genetic determinant(s) detected in their genome; true negatives, if phenotypically susceptible in the absence of any genetic determinant; false positive, if phenotypically susceptible in the presence of a known genetic determinant; and false negative, if phenotypically resistant (or non-susceptible) but not carrying known resistance-conferring genetic determinant(s) in their genome. We calculated commonly used metrics of diagnostic performance using the epi.tests function from the epiR v2.0.43 R package on R v4.1.0, including: negative predictive value (NPV), positive predictive value (PPV), sensitivity and specificity. 95% confidence intervals were calculated using exact binomial confidence limits.

We additionally calculated the diagnostic performance of resistance and non-susceptibility genotypic predictions using only EUCAST clinical breakpoints or only CLSI breakpoints.

Prioritization criteria and AST methods used to re-test isolates with incongruent genotypic predictions

Due to the large number of antibiotics tested (n=15) and isolates analyzed (n=4,730, of which 3,435 from internal collections were available for re-testing), we prioritized a subset of isolates with incongruent genotypic predictions (i.e. false negatives and false positives) for antibiotic susceptibility re-testing to resolve the source of phenotype-genotype discrepancies. No isolates were re-tested for ampicillin, ciprofloxacin and vancomycin as almost perfect phenotype-genotype concordance was achieved for these antibiotics using our curated database. No isolates could be re-tested for daptomycin as all isolates tested for this antibiotic were obtained from published studies (i.e. not available from internal collections). For the remaining antibiotics, we focused on those with unexpectedly high number of false negatives (erythromycin and linezolid), for which phenotypic resistance should have been explained by known mechanisms of resistance; and on antibiotics with a high number of false positives (gentamicin, tetracycline and teicoplanin), i.e. with phenotypically-susceptible isolates carrying known ABR genes. For linezolid we re-tested all 30 false negatives. For erythromycin, due to the large number of false negatives (n=160), we re-tested five randomly selected false negatives clustered in large phylogenetic clades (presumably more likely to carry a novel mechanism of resistance) and five sporadic false negatives (presumably more likely to represent a phenotypic testing error). For gentamicin and tetracycline, antibiotics with large numbers of false positives, we followed similar selection criteria: five false positives clustered in large clades of false positives were selected (presumably more likely to represent cases of ABR gene silencing); and five sporadic false positives surrounded by true positives of the same

or neighboring phylogenetic clade (presumably more likely to represent a phenotypic testing error). For teicoplanin, all false positives from the same study (gouliouris2020, n=12/52) were re-tested. All susceptibility re-testing was performed from frozen stocks, plated onto CBA and grown overnight at 37°C. ETEST®s (biomérieux, France) were performed according to EUCAST guidelines.[61]

DATA ACCESS

The whole genome sequences of the strain collections used in this study are available on European Nucleotide Archive (ENA) under the accessions listed in Supplementary Data 1, which also includes the antibiotic susceptibility phenotypes and genotypes. 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 Francesc Coll. This publication was also supported by the Health Innovation Challenge Fund (WT098600, HICF-T5-342), a parallel funding partnership between the Department of Health and Wellcome Trust. We acknowledge the British Society for Antimicrobial Chemotherapy (BSAC) as a source of *E. faecium* isolates in this study. The Addenbrookes ICU study was supported by a Clinician Scientist Fellowship to MET, funded by the Academy of Medical Sciences and the Health Foundation, and by the NIHR Cambridge Biomedical Research Centre. The Vietnam ICU 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). C.L. was supported by a Wellcome Trust Sir Henry 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 (ESCMID) awarded to K.E.R. L.W.R. was funded by an EBPOD fellowship from EMBL. The views expressed in this publication are those of the author(s) and not necessarily those of the funders.

AUTHOR CONTRIBUTIONS

F. C. and T.G. designed the study with input from all other authors. F.C. undertook the bioinformatic analyses with contributions from T.G. T.G., K.E.R., C.L., B.B., E.M.H., C.S.H., N.M.B., M.A.H., J.P., M.E.T. and S.J.P. all contributed to the generation of the genomic and antibiotic susceptibility data used in this study. H.J.W., F.A.K., and L.W.R. contributed specifically to the laboratory work and genomic analyses for Addenbrookes ICU and Vietnam ICU studies. L.T.H, N.T.H., N.V.T. and N.V.K. contributed to generating the data of the Vietnam ICU study. B.B. conducted the antibiotic susceptibility testing work needed to re-test isolates with incongruent genotypic predictions. F.C., T.G. and S.J.P. wrote the first draft of the manuscript. S.J.P. and J.Pa. supervised the study. All authors had access to the data and read, contributed and approved the final manuscript.

DISCLOSURE DECLARATION

The authors declare no competing interests. SJP and JP are consultants to Next Gen Diagnostics.

FIGURES LEGENDS

Figure 1. MIC distributions with antibiotic resistance determinants included in the curated database. These bar plots show the number of isolates (in the y-axis) with different MICs (x-axis) per antibiotic. 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 showed with different colours; see legend at the right-hand side of each plot. Vertical dotted lines indicate the MIC breakpoints used in this

study (see Supplementary Table 1). For the antibiotic doxycycline, most AST data available was obtained with the disc diffusion method, so no MIC data was available for plotting.

Figure 2 Variants of *vanA* and *vanB* operons found in vancomycin resistant and susceptible isolates. Panels A to C show the variants of *vanA* and *vanB* operons found in vancomycin-resistant isolates (i.e. putatively functional). No *vanB* operon variants were found among vancomycin-resistant isolates, apart from the complete operon. Panel D shows operon variants found in vancomycin-susceptible isolates (i.e. putatively non-functional) that contain the ligase van gene (*vanA* or *vanB*). Other rare operon variants lacking the van gene were also detected (n=7, not shown).

Figure 3. Diagnostic accuracy of genotypic antibiotic resistance predictions of the curated database compared to other databases. This figure shows the sensitivity and specificity of genotypic antibiotic resistance predictions when using the curated database compared to the values obtained by AMRFinder, ResFinder and CARD RGI. See Materials and Methods for details on database versions used. The sensitivity and specificity obtained by LRE-Finder to determine linezolid resistance (81% and 98.8%, respectively) was comparable to those obtain by the curated database (79.6% and 98.2%, respectively).

Figure 1 MIC distributions with antibiotic resistance determinants included in the curated database

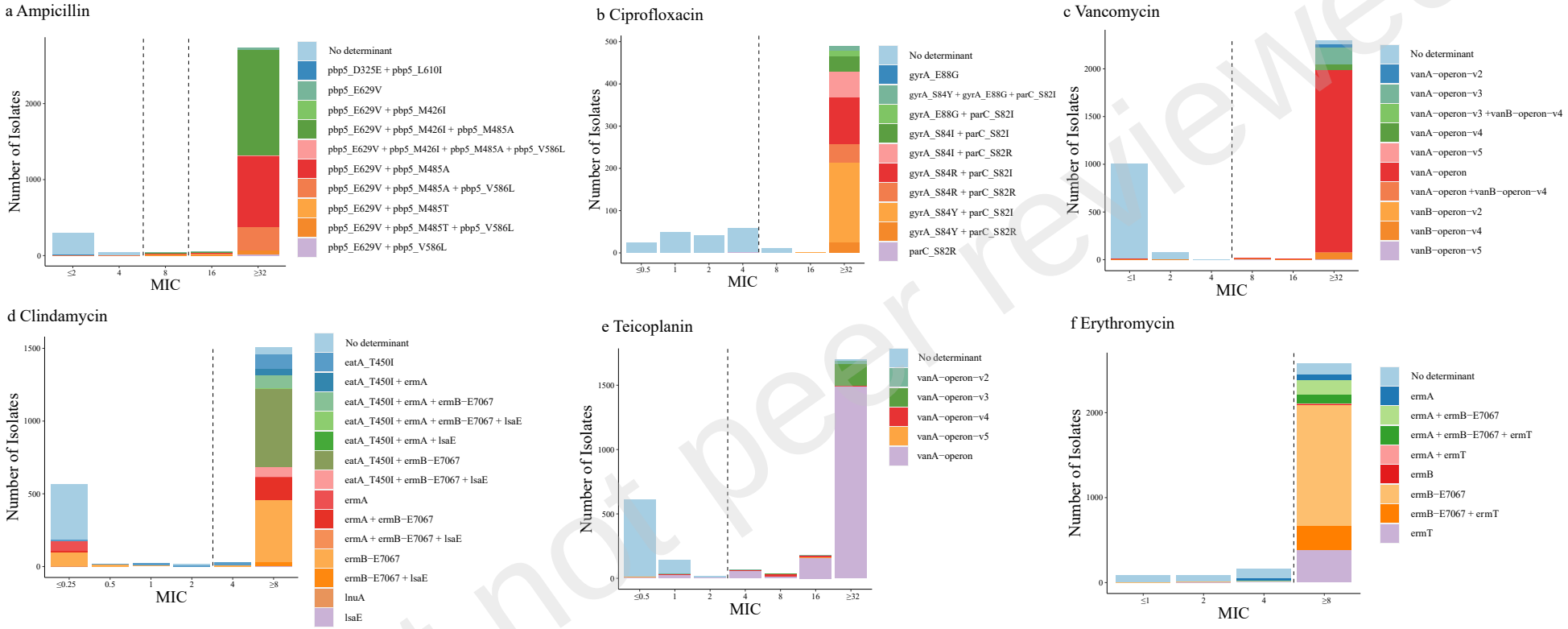


Figure 1 MIC distributions with antibiotic resistance determinants (continued)

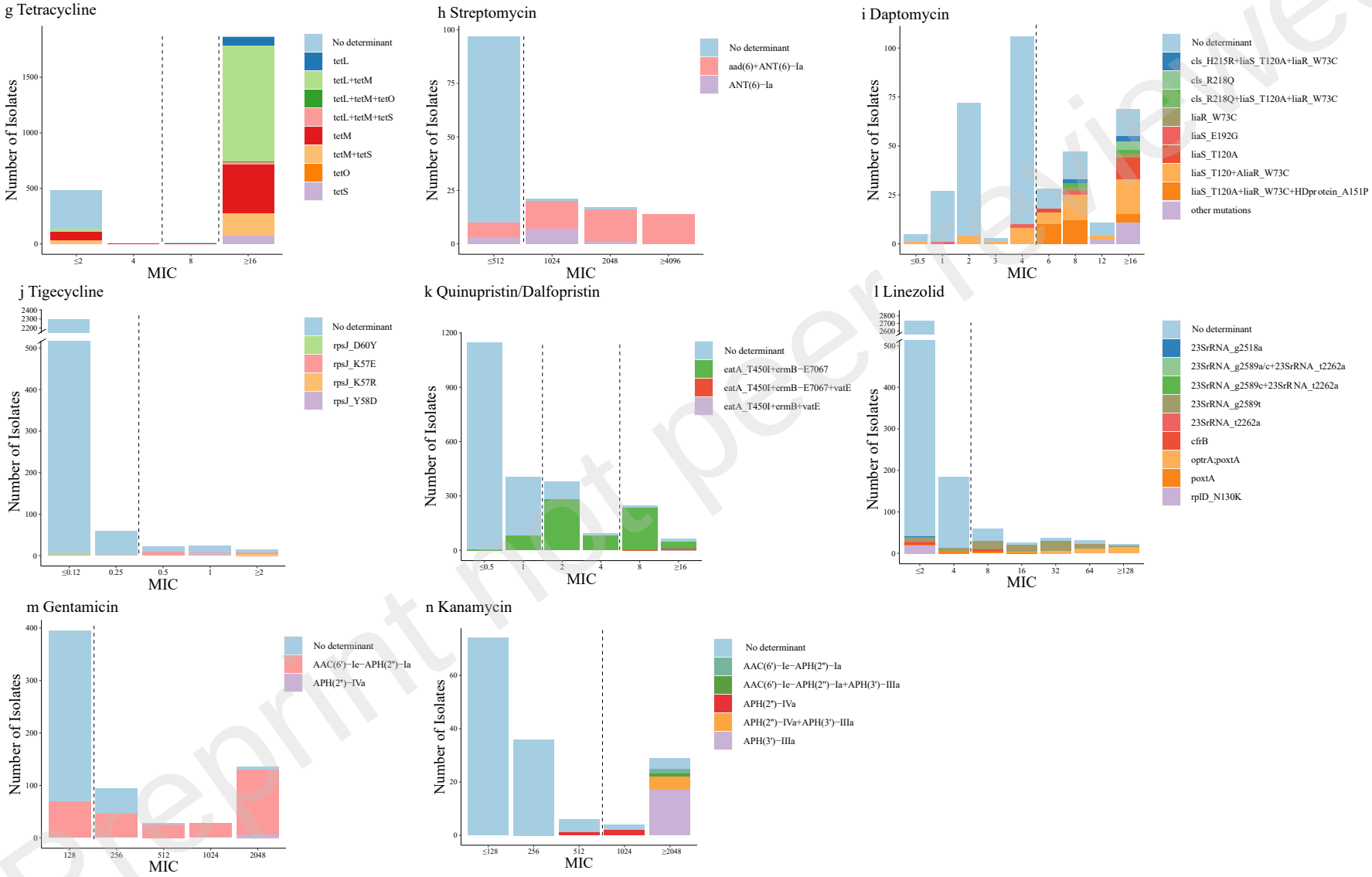


Figure 2 Variants of *vanA* and *vanB* operons found in vancomycin resistant and susceptible isolates

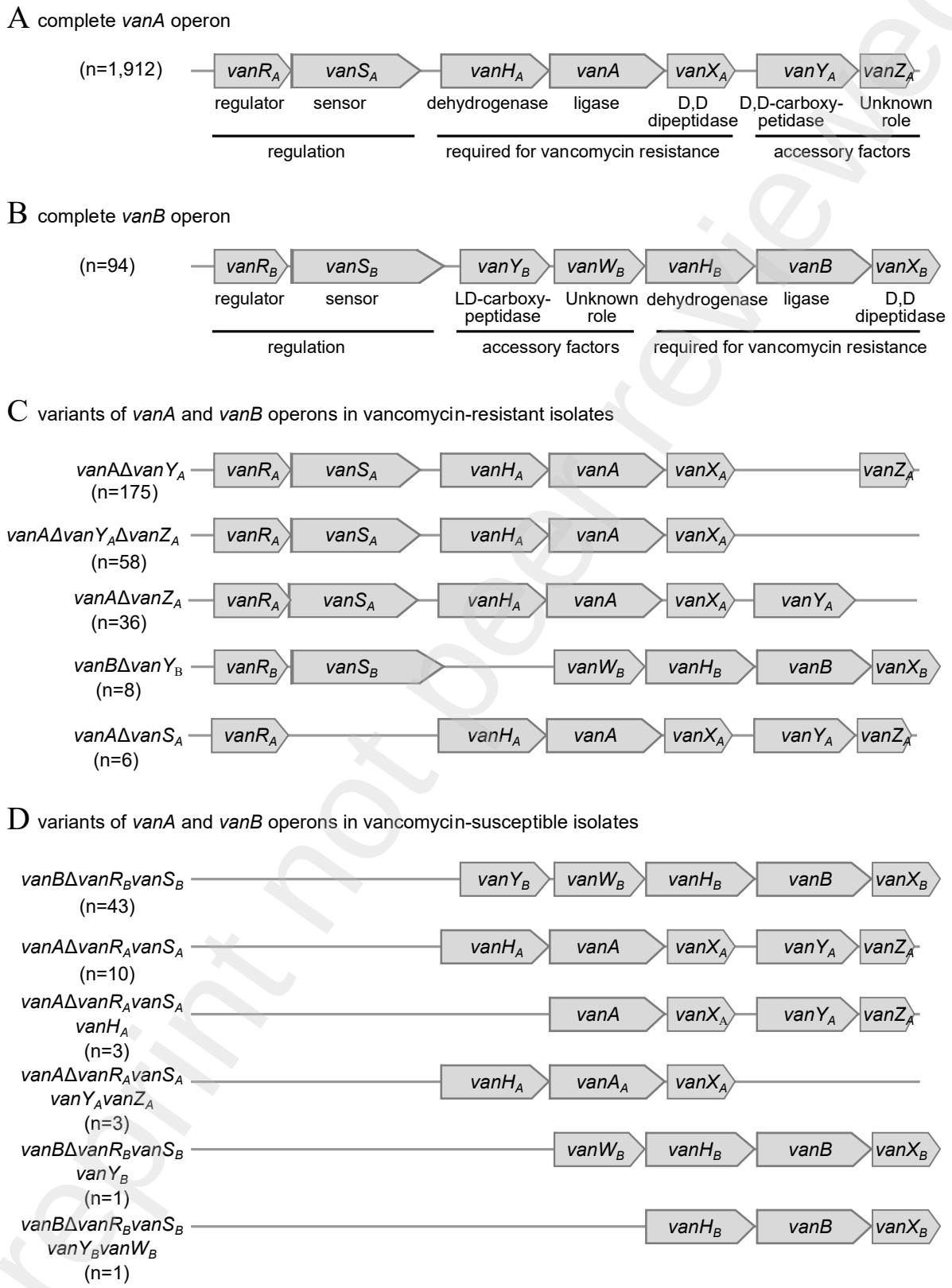
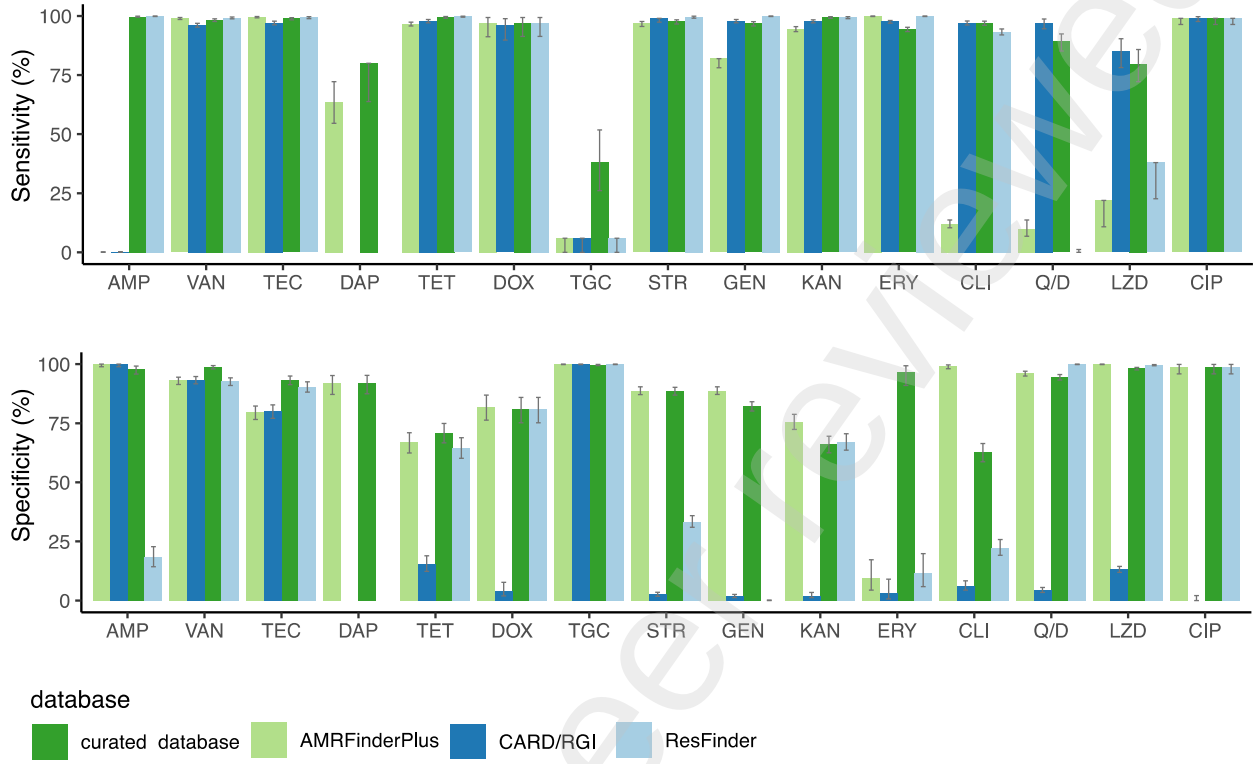


Figure 3 Diagnostic accuracy of genotypic antibiotic resistance predictions of the curated database compared to other databases



786 **Table 1** Characteristics of *E. faecium* collections used in this study

Study Id	# isolates	Antibiotics tested and used	AST methods	Geographical origin (year)	Sources
gouliouris2021[35]	1658	AMP, VAN, TEC, TET, TGC, STR, GEN, KAN, ERY, CLI, Q/D, LZD	Vitek2 P607	UK (2015)	Clinical
raven2016[37]	495	AMP, VAN, TEC, CIP, GEN, ERY, LZD	Vitek2 P607 or ADM	UK (2001–2011)	Clinical
gouliouris2019[38]	398	AMP, VAN, TEC, TET, TGC, STR, GEN, KAN, ERY, CLI, Q/D, LZD	Vitek2 P607	UK (2014-2015)	Wastewater
zaheer2020[20]	336	AMP, VAN, TEC, DOX, TGC, STR, GEN, ERY, Q/D, LZD	Disk diffusion	USA (2014-2016)	Clinical, cattle, beef, water and wastewater
raven2017[36]	293	AMP, VAN, TEC, TET, TGC, STR, GEN, KAN, ERY, CLI, Q/D, LZD	Vitek2 P607	UK (2006-2012)	Clinical
gouliouris2018[34]	267	AMP, VAN, TEC, TET, TGC, STR, GEN, KAN, ERY, CLI, Q/D, LZD	Vitek2 P607	UK (2014-2015)	Chicken, turkey, pig and cattle
Cambridge ICU	243	AMP, VAN, TEC, TET, TGC, STR, GEN, KAN, ERY, CLI, Q/D, LZD	Vitek2 P607	UK (2016)	Clinical
van-hal2016[21]	132	VAN	Vitek2	Australia (2004–2007 and 2011–2013)	Clinical
rushton2019[22]	116	AMP, VAN, TET, ERY	BMD or E-test	New Zealand (2000-2003)	Clinical and poultry
wang2018[23]	106	VAN, DAP	BMD (MicroScan 96Plus) and E-test (DAP)	2009-2013	Clinical
tyson2018[24]	100	VAN, DAP, CIP, TET, TGC, STR, GEN, KAN, ERY, Q/D, LZD	BMD CMV3AGPF	USA (2014)	Cattle, chicken, turkey and pork
wardenburg2019[25]	96	VAN, DAP, LZD	Disk diffusion (VAN) or E-test	USA and Pakistan (2012 - 2018)	Environmental or clinical
Vietnam ICU[60]	81	AMP, VAN, TEC, TET, TGC, STR, GEN, KAN, ERY, CLI, Q/D, LZD	Vitek2 P607	Vietnam (2017-2018)	Clinical
bortolaia2020[18]	50	AMP, VAN, CIP, TET, GEN, ERY, LZD	BMD	Denmark (not reported)	Clinical
zankari2013[4]	49	AMP, VAN, TEC, CIP, TET, TGC,	BMD	Denmark (2011)	Pigs

		STR, GEN, KAN, ERY, LZD			
kerschner2019[26]	46	LZD	E-test	Austria (2014)	Clinical
manson2019[27]	43	VAN, DAP, CIP, TET, TGC, GEN, KAN, ERY, LZD	BMD	USA (2013)	Chicken meat
zhou2018[28]	37	VAN	Vitek2	Netherlands (2014)	Clinical
hasman2019[9]	36	LZD	BMD	Denmark (2015-2018)	Clinical
roach2015[29]	35	AMP, VAN, DAP, CIP, TET, TGC, ERY, LZD	BMD (Sensititre)	USA (2012-2013)	Clinical
douglas2019[30]	31	DAP	E-test	Australia (2014-2015)	Clinical
prater2019[31]	29	AMP, VAN, DAP, ERY	BMD	Not specified	Clinical
kessel2020[32]	28	TGC	Vitek2 and E-test	Germany (2014-2017)	Clinical
kwak2020[33]	25	AMP, VAN, DOX, LZD	Disk diffusion	USA and Canada (2014-2015)	Clinical
Overall	4730	AMP, VAN, TEC, DAP, CIP, TET, DOX, TGC, STR, GEN, KAN, ERY, CLI, Q/D, LZD	Vitek2, Disk diffusion, BMD, BMD (Sensititre), BMD (MicroScan 96Plus) and E-test	UK, Denmark, Netherlands, Germany, Austria, USA, Canada, Australia, New Zealand, Vietnam and Pakistan	Clinical, water, wastewater, chicken, turkey, pigs, cattle, chicken meat, beef and pork

Characteristics and number of *E. faecium* isolates of the 24 collections used in this study. Collections are ordered decreasingly by the number of available *E. faecium* isolates. Isolate-level metadata, including genome accession numbers and AST phenotypic metadata can be found in Supplementary Data 1. Only antibiotics with at least 50 susceptible and 50 resistant isolates across studies were considered and presented in this table. Antibiotic name abbreviations: AMP, ampicillin; VAN, vancomycin; TEC, teicoplanin; DAP, daptomycin; CIP, ciprofloxacin; TET, tetracycline; DOX, doxycycline; TGC, tigecycline; STR, streptomycin; GEN, gentamicin; KAN, kanamycin; ERY, erythromycin; CLI, clindamycin; Q/D, Quinupristin/Dalfopristin; LZD, linezolid. Other abbreviations: AST; antibiotic susceptibility testing; ADM, agar dilution method; BMD, broth microdilution.

Table 2 Summary of antibiotic resistance genetic determinants included in the curated database

797

	single acquired genes		multiple acquired genes		single mutations		multiple mutations	
Antibiotic	#	Gene names	#	Operon or gene names	#	Gene names	#	Gene names
ampicillin	-	-	1	blaZ operon	10	*pbp5* (n=10)	7	*pbp5* (n=7)
vancomycin	-	-	10	vanA operon variants[a] and vanB operon variants[b]	-	-	-	-
teicoplanin	-	-	5	vanA operon variants[a]	-	-	-	-
daptomycin	-	-	-	-	14	*cls* (n=7), *liaF* (n=2), *liaR* (n=2), *liaS* (n=3)	28	*lia* gene+*cls* (n=14), *lia* genes (n=3) and others (n=11)[c]
ciprofloxacin	-	-	-	-	15	*gyrA* (n=12), *parC* (n=3)	30	*gyrA*+*gyrB*+*parC* (n=2), *gyrA*+*parC* (n=28)
tetracycline	8	*tetL*, *tetM*, *tetO*, *tetS*, *tetWNW*, *tetK*, *tetT*, *tetW*	-	-	-	-	-	-
tigecycline	1	*tetX1*	-	-	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*	-	-	-	-	-	-
gentamicin	6	*AAC(6')-Ie-APH(2'')-Ia*, *APH(2'')-IIa*, *APH(2'')-IIIa*, *APH(2'')-IVa*, *APH(6)-Ic*	-	-	-	-	-	-
kanamycin	6	*AAC(6')-Ie-APH(2'')-Ia*, *APH(2'')-IIa*, *APH(2'')-IIIa*, *APH(2'')-IVa*, *APH(3')-IIIa*, *ANT(4')-Ia*	-	-	-	-	-	-
erythromycin	6	*ermB*, *ermB-E7067*, *ermA*, *ermT*, *ermC*, *mefA*	-	-	-	-	-	-
clindamycin	7	*ermB*, *ermB-E7067*, *ermA*, *lsaA*, *lsaE*, *lnuA*, *lnuB*	-	-	1	*eatA* (n=1)	-	-
strep. A[d]	5	*lsaA*, *vgaD*, *vatH*, *vatD*, *vatE*	-	-	1	*eatA* (n=1)	-	-
strep. B[d]	4	*lsaA*, *vgbA*, *ermB*, *ermB-E7067*	-	-	-	-	-	-
Q-D	-	-	1	*vatD* + *vgbA* genes	-	-	-	-
linezolid	4	*cfrA*, *cfrB*, *optrA*, *poxtA*	-	-	20	*23SrRNA* (n=15), *rplC* (n=1), *rplD* (n=3), *rplV* (n=1)	2	*23SrRNA* (n=1), *23SrRNA*+*rplC* (n=1)

798

This table summarises 212 genetic determinants included in the curated database and predictive of antibiotic resistance in *E. faecium*. See Supplementary Data 2 for the complete database. The dash symbol '-' indicates lack of genet *vanA* operon, *vanA*Δ*vanS*$_A$, vanAΔ*vanYA*, vanAΔ*vanZA* and vanAΔ*vanYA*Δ*vanZA*. [b] 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 vanBΔ*vanRB*::*vanRB*-pE7948. [c] Other rare combinations of mutations associated with daptomycin resistance. [d] Quinupristin-Dalfopristin resistance was predicted if isolates carried determinants for both streptogramin A and streptogramin B. Abbreviations: strep. A, streptogramin A; strep. B, streptogramin B; Q-D, quinupristin-dalfopristin. Cross-resistances across fluoroquinolones (i.e. ciprofloxacin), and across tetracyclines (i.e. doxycycline and tetracycline) was assumed.

831 **Table 3** Accuracy of genotypic predictions of phenotypic resistance when using the curated database
832

Antibiotic	Resistant	Susceptible	Not tested	TP	FN	TN	FP	Sensitivity	Specificity	PPV	NPV
ampicillin	2792	340	1599	2785	7	333	7	99.7 (99.5-99.9)	97.9 (95.8-99.2)	99.7 (99.5-99.9)	97.9 (95.8-99.2)
vancomycin	2320	1089	1322	2281	39	1076	13	98.7 (98.1-99.1)	98.8 (98-99.4)	99.4 (99-99.7)	96.5 (95.2-97.5)
teicoplanin	1982	778	1971	1961	21	726	52	98.9 (98.4-99.3)	93.3 (91.3-95)	97.4 (96.6-98.1)	97.2 (95.7-98.3)
daptomycin	127	213	4391	92	35	196	17	72.4 (63.8-80)	92 (87.5-95.3)	84.4 (76.2-90.6)	84.8 (79.6-89.2)
ciprofloxacin	501	172	4058	491	10	170	2	98 (96.4-99)	98.8 (95.9-99.9)	99.6 (98.5-100)	94.4 (90-97.3)
tetracycline	1869	489	2373	1860	9	347	142	99.5 (99.1-99.8)	71 (66.7-74.9)	92.9 (91.7-94)	97.5 (95.3-98.8)
doxycycline	99	221	4411	96	3	179	42	97 (91.4-99.4)	81 (75.2-85.9)	69.6 (61.2-77.1)	98.4 (95.3-99.7)
tigecycline	60	2352	2319	23	37	2346	6	38.3 (26.1-51.8)	99.7 (99.4-99.9)	79.3 (60.3-92)	98.4 (97.9-98.9)
streptomycin	1152	1483	2096	1125	27	1314	169	97.7 (96.6-98.4)	88.6 (86.9-90.2)	86.9 (85-88.7)	98 (97.1-98.7)
gentamicin	1610	1515	1606	1558	52	1246	269	96.8 (95.8-97.6)	82.2 (80.2-84.1)	85.3 (83.6-86.9)	96 (94.8-97)
kanamycin	1924	716	2091	1913	11	473	243	99.4 (99-99.7)	66.1 (62.5-69.5)	88.7 (87.3-90)	97.7 (96-98.9)
erythromycin	2824	95	1812	2664	160	92	3	94.5 (93.6-95.4)	96.8 (91-99.3)	99.9 (99.7-100)	36.5 (30.6-42.8)
clindamycin	1531	623	2577	1486	45	390	233	97.1 (96.1-97.8)	62.6 (58.7-66.4)	86.4 (84.7-88)	89.7 (86.4-92.4)
Q-D	314	1550	2867	280	34	1465	85	89.2 (85.2-92.4)	94.5 (93.3-95.6)	76.7 (72-81)	97.7 (96.8-98.4)
linezolid	147	2903	1681	117	30	2850	53	79.6 (72.2-85.8)	98.2 (97.6-98.6)	68.8 (61.3-75.7)	99 (98.5-99.3)
linezolid*	117	2933	1681	117	0	2880	53	100 (96.9-100)	98.2 (97.6-98.6)	68.8 (61.3-75.7)	100 (99.9-100)

Metrics of diagnostic performance when using the curated database to predict phenotypic antibiotic resistance. Here the intermediate susceptibility category is ignored, i.e. isolates with intermediate susceptibility are treated as not tested. See Supplementary Table 3 for the diagnostic performance achieved by when predicting phenotypic non-susceptibility, that is, when treating the intermediate and resistance categories as non-susceptible. See Supplementary Table 1 and 2 for MIC and zone diameter breakpoints used in this study. Abbreviations: QD, quinupristin-dalfopristin; TP, true positive (phenotypically-resistant isolate carrying an AMR genetic determinant); FN false negative (phenotypically-resistant isolate not carrying any AMR genetic determinant); TN, true negative (phenotypically-susceptible isolate not carrying any AMR genetic determinant); FP, false positive (phenotypically-susceptible isolate carrying an AMR genetic determinant); PPV, positive predictive value; NPV, negative predictive value. *accuracy of genotypic predictions for linezolid after re-testing 30 false negatives, which tested susceptible using E-test.

REFERENCES

1. Murray, C. J. *et al.* Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet* **399**, 629–655 (2022).
2. Ayobami, O., Willrich, N., Reuss, A., Eckmanns, T. & Markwart, R. The ongoing challenge of vancomycin-resistant Enterococcus faecium and Enterococcus faecalis in Europe: an epidemiological analysis of bloodstream infections. *Emerging microbes & infections* **9**, 1180–1193 (2020).
3. Tacconelli, E. *et al.* Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *The Lancet Infectious Diseases* **18**, 318–327 (2018).
4. Zankari, E. *et al.* Genotyping using whole-genome sequencing is a realistic alternative to surveillance based on phenotypic antimicrobial susceptibility testing. *Journal of Antimicrobial Chemotherapy* **68**, 771–777 (2013).
5. Köser, C. U. *et al.* Whole-Genome Sequencing for Rapid Susceptibility Testing of M. tuberculosis. *New England Journal of Medicine* **369**, 290–292 (2013).
6. Stoesser, N. *et al.* Predicting antimicrobial susceptibilities for Escherichia coli and Klebsiella pneumoniae isolates using whole genomic sequence data. *Journal of Antimicrobial Chemotherapy* **68**, 2234–2244 (2013).
7. Feldgarden, M. *et al.* AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Scientific Reports* **11**, 12728 (2021).
8. Florensa, A. F., Kaas, R. S., Clausen, P. T. L. C., Aytan-Aktug, D. & Aarestrup, F. M. ResFinder - an open online resource for identification of antimicrobial resistance genes in next-generation sequencing data and prediction of phenotypes from genotypes. *Microbial genomics* **8**, 1–10 (2022).

9. Hasman, 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. *Journal of Antimicrobial Chemotherapy* **74**, 1473–1476 (2019).

10. Hunt, M. *et al.* Antibiotic resistance prediction for Mycobacterium tuberculosis from genome sequence data with Mykrobe. *Wellcome open research* **4**, 191 (2019).

11. Coll, F. *et al.* Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. *Genome Medicine* **7**, 51 (2015).

12. Walker, A. S. & Peto, T. E. A. Prediction of Susceptibility to First-Line Tuberculosis Drugs by DNA Sequencing. *New England Journal of Medicine* **379**, 1403–1415 (2018).

13. Phelan, J. E. *et al.* Integrating informatics tools and portable sequencing technology for rapid detection of resistance to anti-tuberculous drugs. *Genome Medicine* **11**, 41 (2019).

14. Bradley, P. *et al.* Rapid antibiotic-resistance predictions from genome sequence data for Staphylococcus aureus and Mycobacterium tuberculosis. *Nature Communications* **6**, 10063 (2015).

15. Gladstone, R. A. *et al.* International genomic definition of pneumococcal lineages, to contextualise disease, antibiotic resistance and vaccine impact. *EBioMedicine* **43**, 338–346 (2019).

16. Anahtar, M. N. *et al.* Prediction of Antimicrobial Resistance in Clinical Enterococcus faecium Isolates Using a Rules-Based Analysis of Whole-Genome Sequences. *Antimicrobial Agents and Chemotherapy* **66**, AAC0119621 (2022).

17. Tyson, G. H., Sabo, J. L., Rice-Trujillo, C., Hernandez, J. & McDermott, P. F. Whole-genome sequencing based characterization of antimicrobial resistance in Enterococcus.

Pathogens and Disease **76**, 1–5 (2018).

18. Bortolaia, V. *et al.* ResFinder 4.0 for predictions of phenotypes from genotypes. *Journal of Antimicrobial Chemotherapy* **75**, 3491–3500 (2020).

19. Penven, M. *et al.* Web-based prediction of antimicrobial resistance in enterococcal clinical isolates by whole-genome sequencing. *European Journal of Clinical Microbiology & Infectious Diseases* (2022) doi:10.1007/s10096-022-04527-z.

20. Zaheer, R. *et al.* Surveillance of Enterococcus spp. reveals distinct species and antimicrobial resistance diversity across a One-Health continuum. *Scientific Reports* **10**, 3937 (2020).

21. van Hal, S. J. *et al.* Evolutionary dynamics of Enterococcus faecium reveals complex genomic relationships between isolates with independent emergence of vancomycin resistance. *Microbial Genomics* **2**, 1–11 (2016).

22. Rushton-Green, R. *et al.* Agricultural Origins of a Highly Persistent Lineage of Vancomycin-Resistant Enterococcus faecalis in New Zealand. *Applied and Environmental Microbiology* **85**, 1–17 (2019).

23. Wang, G. *et al.* Evolution and mutations predisposing to daptomycin resistance in vancomycin-resistant Enterococcus faecium ST736 strains. *PLOS ONE* **13**, e0209785 (2018).

24. Tyson, G. H., Sabo, J. L., Rice-Trujillo, C., Hernandez, J. & McDermott, P. F. Whole-genome sequencing based characterization of antimicrobial resistance in Enterococcus. *Pathogens and Disease* **76**, 1–5 (2018).

25. Wardenburg, K. E. *et al.* Phenotypic and genotypic characterization of linezolid-resistant Enterococcus faecium from the USA and Pakistan. *Journal of Antimicrobial Chemotherapy* **74**, 3445–3452 (2019).

26. Kerschner, H. *et al.* Hospital outbreak caused by linezolid resistant Enterococcus

faecium in Upper Austria. *Antimicrobial Resistance & Infection Control* **8**, 150 (2019).

27. Manson, A. L. *et al.* Chicken Meat-Associated Enterococci: Influence of Agricultural Antibiotic Use and Connection to the Clinic. *Applied and Environmental Microbiology* **85**, 1–14 (2019).

28. Zhou, X. *et al.* Elucidating vancomycin-resistant Enterococcus faecium outbreaks: the role of clonal spread and movement of mobile genetic elements. *Journal of Antimicrobial Chemotherapy* **73**, 3259–3267 (2018).

29. Roach, D. J. *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 Genetics* **11**, e1005413 (2015).

30. Douglas, A. P. *et al.* Utilizing genomic analyses to investigate the first outbreak of vanA vancomycin-resistant enterococcus in Australia with emergence of daptomycin non-susceptibility. *Journal of Medical Microbiology* **68**, 303–308 (2019).

31. Prater, A. G. *et al.* Environment Shapes the Accessible Daptomycin Resistance Mechanisms in Enterococcus faecium. *Antimicrobial Agents and Chemotherapy* **63**, 3–5 (2019).

32. Kessel, J. *et al.* Risk factors and outcomes associated with the carriage of tigecycline- and vancomycin-resistant. *Journal of Infection* **82**, 227–234 (2021).

33. Kwak, S. *et al.* Impact of investigational microbiota therapeutic RBX2660 on the gut microbiome and resistome revealed by a placebo-controlled clinical trial. *Microbiome* **8**, 125 (2020).

34. Gouliouris, T. *et al.* Genomic Surveillance of Enterococcus faecium Reveals Limited Sharing of Strains and Resistance Genes between Livestock and Humans in the United Kingdom. *mBio* **9**, 1–15 (2018).

35. Gouliouris, T. *et al.* Quantifying acquisition and transmission of Enterococcus faecium

using genomic surveillance. *Nature Microbiology* **6**, 103–111 (2021).

36. Raven, K. E. *et al.* Complex Routes of Nosocomial Vancomycin-Resistant Enterococcus faecium Transmission Revealed by Genome Sequencing. *Clinical Infectious Diseases* **64**, 886–893 (2017).

37. Raven, K. E. *et al.* A decade of genomic history for healthcare-associated Enterococcus faecium in the United Kingdom and Ireland. *Genome Research* **26**, 1388–1396 (2016).

38. Gouliouris, T. *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 Research* **29**, 626–634 (2019).

39. Rice, L. B. *et al.* Impact of Specific pbp5 Mutations on Expression of β-Lactam Resistance in Enterococcus faecium. *Antimicrobial Agents and Chemotherapy* **48**, 3028–3032 (2004).

40. Gagnon, S. *et al.* vanA-containing Enterococcus faecium susceptible to vancomycin and teicoplanin because of major nucleotide deletions in Tn1546. *Journal of Antimicrobial Chemotherapy* **66**, 2758–2762 (2011).

41. Park, I. J., Lee, W. G., Shin, J. H., Lee, K. W. & Woo, G. J. VanB Phenotype- vanA Genotype Enterococcus faecium with Heterogeneous Expression of Teicoplanin Resistance. *Journal of Clinical Microbiology* **46**, 3091–3093 (2008).

42. Hashimoto, Y., Tanimoto, K., Ozawa, Y., Murata, T. & Ike, Y. Amino acid substitutions in the VanS sensor of the VanA-type vancomycin-resistant enterococcus strains result in high-level vancomycin resistance and low-level teicoplanin resistance. *FEMS Microbiology Letters* **185**, 247–254 (2000).

43. Lauderdale, T.-L. *et al.* Vancomycin-Resistant Enterococci from Humans and Retail Chickens in Taiwan with Unique VanB Phenotype- vanA Genotype Incongruence.

Antimicrobial Agents and Chemotherapy **46**, 525–527 (2002).

44. Alcock, B. P. *et al.* CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Research* **48**, 517–525 (2019).

45. Portillo, A. *et al.* Macrolide resistance genes in Enterococcus spp. *Antimicrobial Agents and Chemotherapy* **44**, 967–971 (2000).

46. Pietta, E., Montealegre, M. C., Roh, J. H., Cocconcelli, P. S. & Murray, B. E. Enterococcus faecium PBP5-S/R, the Missing Link between PBP5-S and PBP5-R. *Antimicrobial Agents and Chemotherapy* **58**, 6978–6981 (2014).

47. Kime, L. *et al.* Transient silencing of antibiotic resistance by mutation represents a significant potential source of unanticipated therapeutic failure. *mBio* **10**, (2019).

48. Heyckendorf, J. *et al.* What Is Resistance? Impact of Phenotypic versus Molecular Drug Resistance Testing on Therapy for Multi- and Extensively Drug-Resistant Tuberculosis. *Antimicrobial Agents and Chemotherapy* **62**, e01550-17 (2018).

49. 'Geneva: World Health Organization'. *Technical Report on critical concentrations for drug susceptibility testing of medicines used in the treatment of drug-resistant tuberculosis*. (2018).

50. Chow, J. W. Aminoglycoside Resistance in Enterococci. *Clinical Infectious Diseases* **31**, 586–589 (2000).

51. Miller, W. R., Munita, J. M. & Arias, C. A. Mechanisms of antibiotic resistance in enterococci. *Expert Review of Anti-Infective Therapy* **12**, 1221–1236 (2014).

52. Bender, J. K. *et al.* Update on prevalence and mechanisms of resistance to linezolid, tigecycline and daptomycin in enterococci in Europe: Towards a common nomenclature. *Drug Resistance Updates* **40**, 25–39 (2018).

53. Hershberger, E., Donabedian, S., Konstantinou, K. & Zervos, M. J. Quinupristin-

Dalfopristin Resistance in Gram-Positive Bacteria: Mechanism of Resistance and Epidemiology. *Clinical Infectious Diseases* **38**, 92–98 (2004).

54. Hollenbeck, B. L. & Rice, L. B. Intrinsic and acquired resistance mechanisms in enterococcus. *Virulence* **3**, 421–569 (2012).

55. Kristich, C. J.; Rice, L. B. . A. C. A. Enterococcal Infection− Treatment and Antibiotic Resistance. In Enterococci: From Commensals to Leading Causes of Drug Resistant Infection. 123–185 (2014).

56. Gagetti, P. *et al.* Resistance to β-lactams in enterococci. *Revista Argentina de Microbiologia* **51**, 179–183 (2019).

57. Raven, K. E. *et al.* Methodology for Whole-Genome Sequencing of Methicillin-Resistant *Staphylococcus aureus* Isolates in a Routine Hospital Microbiology Laboratory. *Journal of Clinical Microbiology* **57**, (2019).

58. Bortolaia, V. *et al.* ResFinder 4.0 for predictions of phenotypes from genotypes. *Journal of Antimicrobial Chemotherapy* **75**, 3491–3500 (2020).

59. Zankari, E. *et al.* PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. *Journal of Antimicrobial Chemotherapy* **72**, 2764–2768 (2017).

60. Roberts, L. W. *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. *The Lancet Microbe* **3**, e857–e866 (2022).

61. EUCAST. Antimicrobial wild type distributions of microorganisms. https://mic.eucast.org/search/ (2022).

62. EUCAST. Clinical breakpoints and dosing of antibiotics. v12.0 https://www.eucast.org/clinical_breakpoints/ (2022).

63. CLSI. CLSI M100-ED32:2022 Performance Standards for Antimicrobial Susceptibility Testing, 32nd Edition. http://em100.edaptivedocs.net/ (2022).
64. Hunt, M. *et al.* ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microbial Genomics* **3**, e000131 (2017).