Molecular characterization of vancomycin-resistant *Enterococcus faecium* isolates collected from the Dallas, Texas area

Moutusee Islam,^a Ada Abaragu,^a Harita Sistu,^{a*} Ronda L. Akins^b and Kelli Palmer^{a#}

^aDepartment of Biological Sciences, University of Texas at Dallas, Richardson, Texas, USA

^bMethodist Health System, Dallas, Texas, USA

Running Head: Molecular characterization of Dallas VREfm isolates

#Address correspondence to Kelli Palmer, kelli.palmer@utdallas.edu

*Present address: Harita Sistu, Department of Microbiology, University of Massachusetts Amherst, Massachusetts, USA

Keywords: *Enterococcus*, mobile genetic element, plasmid, vancomycin resistance.

Abstract word count: 238

Importance word count: 130

Text word count: 6,340

Abstract

Molecular characterization was performed for 46 vancomycin-resistant *E. faecium* (VREfm) isolates and one vancomycin-sensitive comparator obtained during routine fecal surveillance of high-risk patients from a Dallas, Texas area hospital system. Hybrid assemblies of long (Oxford Nanopore Technology) and short (Illumina) sequence reads enabled the generation of 31 complete and 16 draft high quality genome sequences. The VREfm isolates possessed up to 12 plasmids each. A total of 251 closed plasmid sequences, assigned to 12 previously described and 9 newly defined *rep* family types, were obtained. Phylogenetic analysis clustered the Dallas isolates into genomic Clade A1, including 10 different sequence types (STs), the most prevalent ones being ST17 and ST18. VREfm isolates with the novel sequence type ST1703 were also identified. All but three of the VREfm isolates encoded *vanA*-type vancomycin resistance within Tn1546-like elements on a pRUM-like plasmid backbone. New variants of the Tn1546 were described that harbored a combination of 7 insertion sequences (IS) including 3 novel IS elements reported in this study (ISEfa16, ISEfa17 and ISEfa18). The pRUM-like plasmids carrying the Tn1546 were grouped into 4 sub-groups based on the replication and stability modules of the plasmid backbone. Overall, we conclude that the VREfm isolates analyzed in our collection are representative of other global VanA-type VREfm in that they belong to the Clade A1 lineage, but they possess novel arrangements of the Tn1546-like elements and the *vanA* operon, which have evolved independently of the pRUM-like backbone.

Importance

Vancomycin is a broad-spectrum antibiotic used to treat infections caused by multi-drug resistant Gram-positive bacteria. However, vancomycin resistance is predominant for *Enterococcus faecium* and therefore infections with VREfm have few treatment options. The vancomycin resistance genes are encoded on a transposon, which can be disseminated by different plasmids with different host ranges and transferring efficiencies. The presence of resistance genes within a broad host range plasmid can generate a wide variety of resistant bacteria very quickly. Therefore, surveillance of resistance plasmids is critical. This study provided a comprehensive analysis of VREfm collected from the Dallas, Texas area, with particular focus on the mobile elements associated with vancomycin resistance genes. We find that a single plasmid family, the pRUM-like family, is associated with vancomycin resistance in the majority of isolates sampled.

Introduction

Originally described as a commensal of the healthy human gut, the enterococci attracted attention in the 1990s due to their alarming multi-drug resistance¹. Increasing cases of vancomycin-resistant enterococci (VRE), first reported in 1988 in Europe² and then in USA in 1989³ became a major concern for nosocomial infection⁴. In recent years, vancomycin-resistant *E. faecium* (VREfm) is considered to be the most prominent VRE species, and hence is widely studied⁴.

Whole genome-based phylogenetic analysis and multi-locus sequence typing (MLST) of *E. faecium* isolates has provided valuable information regarding their origin and evolution. Phylogenetic analysis of *E. faecium* strains collected from sources including healthy human volunteers, animals, and hospital infections revealed the presence of at least two major clades, referred to as Clade A and B, within this species^{5, 6}. Clade B *E. faecium* strains are mostly commensals of the healthy human gut. Clade A strains adapted to colonize the animal gut approximately 3000 years ago, coincident with human urbanization, hygienic behavior, and animal domestication⁵. Clade A *E. faecium* strains diverged further into Clades A1 and A2 around 75 years ago, coinciding with widespread usage of antibiotics⁵. The multi-drug resistant *E. faecium* strains, responsible for hospital outbreaks, typically belong to Clade A1, whereas animal gut commensals typically belong to Clade A2⁵.

MLST analysis, which pre-dates whole genome-based phylogenetic analyses, classifies VREfm isolates by nucleotide sequence variations occurring in 7 housekeeping genes (*adk*, *ddl*, *gdh*, *purK*, *pstS*, *atpA*, and *gyd*)⁷. Isolates are assigned to different sequence types (ST) based on the allelic variants. Most of the reported hospital outbreaks of VREfm emerged from a single genetic lineage of clonal

complex-17 (CC17) that was founded by the ST17⁸. The VREfm isolates involved in hospital outbreaks in different parts of the world, within the time period 1986-2012, represented 26 STs⁹. The most prominent VREfm isolates reported were ST17 (reported in Spain, China, Brazil), ST18 (reported in Canada, Poland, South America, Turkey, Serbia) and ST78 (reported in Turkey, Paraguay, Belgrade, Italy). In USA, VREfm of ST16 and ST17 were responsible for clonal outbreaks from 1990 to 2005. Generally, a single ST of VREfm is associated with an outbreak in a certain geographic region, while globally, outbreaks are caused by multiple different STs. For example, ST25 (France, 1987), ST114 (Brazil, 1998), ST18 (UK, 1992) and ST16 (Netherland, 2002) were reported outbreak strains from different parts of the world⁹. Epidemiological differences in diverse geographical areas⁹ reiterate the importance of tracking VREfm genotypes.

Vancomycin resistance genes are encoded within mobile elements in VREfm, and can be horizontally disseminated¹⁰. The most commonly occurring vancomycin resistance type worldwide, currently, is VanA-type resistance, which is typically encoded within the element Tn1546¹¹. Broadly speaking, the Tn1546 may be chromosomally integrated or plasmid-borne. Different plasmid families carrying Tn1546 including Inc18, pRUM, and pLG1 have been reported worldwide. Hybrids of pRUM-Inc18, Inc18-pHT β , pRUM-pLG1 family plasmids carrying *vanA* have been reported as well⁹. Structural variations of Tn1546 due to the presence of insertion sequences (IS) like IS1251 (USA, Australia, Europe), IS1216 (Europe), IS1485 (Saudi Arabia) and ISEf1 (Germany, Portugal, Poland) have also been reported. Previous studies determined the Tn1546 structural variation by PCR mapping and then sequencing of the overlapping fragments^{9, 12, 13}. However, in the absence of a completely closed genome assembly, conclusive links between *vanA*, Tn1546

variants, and specific plasmid backbones in VREfm is difficult to achieve. Surveillance of resistance elements is important because changes in host range and transfer frequency can evolve, which could impact clinical care. Utilization of whole genome sequencing for epidemiologic study provides an opportunity for multilevel molecular analysis that includes detection of novel mobile genetic elements (MGE) and monitoring changes in resistance gene vectors. The advent of long-read next generation sequencing (NGS) by Oxford Nanopore Technology (ONT) made it possible to resolve long repeat regions and generate genome assemblies of a large number of bacterial isolates¹⁴ in a cost effective manner. The combination of long (ONT) and short (Illumina) reads generates high quality completely closed genome assemblies that gives access to fully assembled mobile genetic elements¹⁵.

Much of the focus of VREfm genomics has been on infection-derived isolates. There is a comparative lack of information regarding genotypic diversity of fecal (i.e., non-infection-derived) VREfm isolates that colonize hospitalized patients in a given population. These represent reservoirs from which future hospital outbreaks could emerge. The current study aims to address this knowledge gap by analyzing fecal VREfm isolates collected from hospitalized patients in Dallas, Texas, using a combination of ONT and Illumina sequencing.

Materials and Methods

Bacterial strains analyzed in this study. The isolates used in this study (**Dataset S1A**) were cultured from surveillance rectal swabs of high-risk patients on hospital admission within a span of two months (August 2015 to October 2015), and their clinical collection and initial characterization was previously reported¹⁶. Briefly, rectal swabs from patients admitted with at least one of the following risk factors

(hospitalization for ≥ 2 consecutive nights in the preceding 30 days, transferred from another medical facility, residence in a nursing home or extended/long term care facility, or the presence of decubitus ulcer or a draining wound) were cultured on Spectra VRE (Remel) agar as a part of routine infection control surveillance. Instead of disposal after clinical procedures were complete, cultures positive for presumptive VRE growth were coded numerically, de-identified, and transferred to the University of Texas at Dallas for further analysis. Isolates were coded by de-identified patient number. If multiple colonies were collected from a Spectra VRE plate, this was indicated by hyphenated numerics. For example, the VRE isolates 124-1, 124-2 and 124-3 were collected from the same patient (de-identified patient #124). In our previous study of these isolates, presumptive VREfm were confirmed for species by *ddl* typing and for the presence of *vanA* or *vanB* genes by PCR^{16,17}. All isolates used in this study, with the exception of 163-1, grew on brain heart infusion (BHI) agar supplemented with 256 $\mu\text{g/mL}$ vancomycin. The vancomycin MIC of isolate 163-1 was previously determined to be 2 $\mu\text{g/mL}$ by broth microdilution¹⁶.

Genome sequencing and assembly. Methods for DNA isolation, MinION sequencing, Illumina sequencing, and hybrid genome assembly are described in the Supplemental Text.

Phylogenetic analyses and MLST. Genome assemblies were annotated with PROKKA (v 1.12)¹⁸. A sequence alignment was generated for 1706 core genes by ROARY (v 3.12)¹⁹ using PRANK where the paralogs were not split. A phylogenetic tree was built by RAxML (v 8.2.12)²⁰ with rapid bootstrapping (1000 inferences) using the GTR+GAMMA model and subsequent ML search on the core gene alignment

where the outgroup was specified to be the reference genome *E. faecium* 1,231,501 (GenBank accession number ACAY01). The phylogenetic tree was visualized with iTOL (v4)²¹. MLST was determined by the *E. faecium* MLST database (<https://pubmlst.org/efaecium/>)^{22, 7}. The novel sequence type ST1703 and novel allele number for *gyd* (allele-70) was assigned for 5 VREfm isolates (111, 121, 137, 154-1, 158) upon request from the database.

Plasmid detection and typing. Each contig of <2 Mbp in size in a closed genome assembly was considered a putative plasmid. *In silico* detection and typing of plasmids was performed with PlasmidFinder²³ (v 2.1). Putative plasmids in closed genome assemblies that were not classifiable by PlasmidFinder were analyzed further (see below).

Definition of pMIX plasmid groups. Putative plasmid sequences that were unclassifiable by PlasmidFinder were analyzed as follows. Presumptive *rep* genes were identified from plasmid annotations and analyzed by NCBI Conserved Domain Analysis (CDS). Confirmed *rep* genes encoded protein domains that matched known Rep protein families with e-values of e^{-10} to e^{-78} . The Rep sequences were also scanned by HMMER (v 2.41.1)²⁴ to analyze pfam domains. Multiple sequence alignment of the nucleotide and predicted amino acid sequences of *rep* genes were performed with MUSCLE (v 3.8.424) where the “group sequences by similarity” option was selected. A neighbor-joining tree was built based on the sequence alignment using the “Tamura Nei” genetic distance model²⁵. The tree was visually scrutinized in order to sort the plasmids into 9 groups harboring *rep* genes with at least 95% identity in both nucleotide and predicted amino acid sequence (except for pMI4; see Results).

Each unique *rep* group of plasmids was given a name of format pMIX, where the “X” is a numeric number from 1 to 9.

Eight closed circular DNA elements of varying sizes (3 to 72 kb) could not be classified by PlasmidFinder nor by the pMIX *rep* typing scheme described above. Analysis of these elements is described in the Supplemental Text.

Transcriptional activity of predicted toxin-antitoxin (TA) systems. Transcriptional activity of the TA systems, $TA_{axe-txe}$ and TA_{relE} , was determined by RT-qPCR for the VREfm isolates 1 and 5, described in the Supplemental Text. The primer sequences used for qPCR are provided in **Dataset S1B**.

Resistance gene identification. Acquired antimicrobial resistance genes were detected with ResFinder 3.1²⁶.

Tn1546 analysis. The Tn1546 nucleotide sequence described in *E. faecium* BM4147 (accession number M97297) was used as a reference to identify variations occurring in the Tn1546 identified in this study. The Tn1546 were classified using a previously described nomenclature¹³. The presence of insertion sequence elements (IS elements) was denoted with a one-letter code, shown in parentheses: IS1216 (B), IS1251 (C), ISEfa5 (D), IS256 (J), ISEfa16 (K), ISEfa17 (L), ISEfa18 (M). The combination of IS elements within the transposon is described by a two or three letter code; i.e., group ‘BC’ possesses one each of IS1216 and IS1251. Arabic numerals, following the alphabet code, indicate differences due to point mutations or different IS element insertion sites. For the group ‘BC’, a previous study¹³ described BC1-BC5, each with specific insertion sites of IS elements. Any novel insertion sites identified in

this present study were therefore numbered from 'BC6' forward. The novel *ISEfa16*, *ISEfa17*, *ISEfa18* sequences were identified in this study and registered in the ISFinder public database²⁷.

Accession numbers. DNA sequences from this study have been deposited under Bioproject PRJNA682584. Individual accession numbers for all sequence files are provided in **Dataset S1A**.

Results

VREfm isolates belonging to ST17 and ST18 were prevalent in the collection.

Genome sequencing of 46 VREfm isolates and one vancomycin-sensitive comparator collected from rectal surveillance of high-risk patients at a Dallas, Texas, hospital was performed with Oxford Nanopore MinION and Illumina technologies. High quality genome sequences were generated by hybrid assembly of the sequencing reads from these two platforms. 31 closed and 16 draft genome assemblies were generated (**Table S1**). In our previous analysis of these isolates using PCR, all isolates were found to be *vanA*-positive and *vanB*-negative by PCR, with the exception of the isolate 163-1, which had neither *vanA* or *vanB* and had a vancomycin MIC of 2 µg/mL by broth microdilution¹⁶. We included 163-1 in this study for comparative purposes to evaluate its relationship to the *vanA*-positive isolates in the collection. All 46 *vanA*-positive isolates sequenced in this study grew with 256 µg/mL vancomycin in the growth medium.

Phylogenetic analysis based on alignment of 1706 core genes allowed us to assess the epidemiology and diversity of the VREfm isolates (**Figure 1**). Two reference

genomes, *E. faecium* 1,231,501⁶ and *E. faecium* ATCC® 700221™²⁸, were included as representatives of *E. faecium* clade A2 and A1, respectively. *E. faecium* 1,231,501 is a human bloodstream isolate from the USA and is susceptible to vancomycin. *E. faecium* ATCC® 700221™ is a USA isolate and VanA-type VREfm control strain used in antimicrobial susceptibility testing²⁹. As expected, the Dallas VREfm isolates clustered with the clade A1 reference *E. faecium* ATCC® 700221™, although a high degree of branching within this clade was observed, revealing genetic diversity within the collection.

MLST is commonly used to study relationships among VREfm isolates worldwide, and this analysis provides the opportunity to compare the Dallas isolates with other VREfm strain collections. A total of 10 different sequence types (ST), all belonging to clonal complex 17 (CC17), were observed among the collection (**Figure 1, Figure 2**). Isolates belonging to ST17 and ST18 are the most prevalent in the collection, comprising 50% of the isolates, which clustered together in the core genome tree (**Figure 1, Figure 2**). Six genotypes (ST18, ST664, ST1703, ST736, ST80 and ST612) are double locus variants of ST17. The ST262 and ST412 are triple locus variants and ST1383 is a single locus variant of ST17 (**Dataset S1C**). ST1703, a double locus variant of ST17 (*pstS* allele-20; novel *gyd* allele-70), is a novel ST in the MLST database and comprised 11% of the collection. Isolate 163-1 branched away from the rest of the isolates and belonged to ST262, which is still within CC17. Taken together, these data indicate that *E. faecium* CC17 with varied genotypes colonized Dallas patients at the time of the study.

Large numbers of plasmids are present in the Dallas isolates. The plasmid content of the Dallas isolates was analyzed by an *in silico* detection method utilizing

the conserved nucleotide sequences of replication initiation (*rep*) genes²³. Of the 259 closed circular putative plasmids, 141 were assigned to 10 known *rep* families described in the existing PlasmidFinder database. The remaining 118 putative plasmids were categorized into 9 novel plasmid groups (pMI1 to pMI9) and 8 other elements (discussed in the next section).

The Dallas isolates harbored a variable number of plasmids (n=3 to 12) of sizes ranging from 1.9 to 294 kb (**Figure 1; Dataset S1D**). The most common elements were a megaplasmid (126 to 294 kb) and a pRUM-like plasmid (17 to 63 kb), which were each present in all isolates except 144-1, which lacked the megaplasmid. Both the reference strains (*E. faecium* 1,231,501 and ATCC® 700221™) harbored the megaplasmid, but 1,231,501 lacked the pRUM-like plasmid. Each of the megaplasms harbored the *rep* gene *repUS15*, encoding a RepA_N protein motif. The pRUM-like plasmid harbored *rep17* or both *rep17* and *rep18*, which we refer to as a mosaic plasmid. Isolates collected from the same patient had identical plasmid profiles in 2 cases (patients 124 and 9) while different plasmid profiles were observed in 3 other cases (patients 100, 71, and 16).

To determine whether plasmid replicons were enriched among specific lineages, the percentage presence/absence of specific *rep* genes was quantified among the isolates belonging to 4 STs (ST17, ST18, ST1703 and ST664) that included ≥5 isolates. The other 6 STs (ST262, ST1383, ST612, ST412, ST80, and ST736) were not included in this analysis because they were represented by <5 isolates (**Dataset S1E**). Chromosomally integrated *rep* genes *repUS12* and *repUS43*, encoding the protein motifs Rep_1 and Rep_trans, were observed in 45% and 22% of the isolates, respectively, and were absent from isolates belonging to ST17, ST18 and ST1383

(**Dataset S1E**). The *rep14* family was present in 80% and 90% of ST17 and ST18 isolates, respectively, but was absent in ST1703 and ST664 isolates. The *rep18* family plasmid was present in ST17 (80%) and ST664 (100%) isolates but absent in ST1703 (0%) and rare in ST18 (30%). In summary, the *rep* genes *repUS12*, *repUS43*, *rep18* and *rep14* had an ST-specific distribution among the isolates. No conclusion could be drawn regarding other *rep* genes and STs due to low sample size.

Uncategorized plasmids were sorted into 9 plasmid groups, pMI1 to pMI9. PlasmidFinder²³ did not identify *rep* genes in a total of 118 circular entities. Of these 118 circular entities, 110 were established to be plasmids based on the presence of confirmed *rep* genes not currently represented in the PlasmidFinder database (see Materials and Methods). The remaining 8 circular entities were designated as putative excision products of genomic islands (**Supplementary text, Dataset S1F-G**) including a 5 kb *vanA*-encoding circular fragment of Tn1546 in isolate 111 (discussed further in a later section).

The 110 plasmids not classifiable by PlasmidFinder were sorted into 9 plasmid groups, pMI1 to pMI9, based on phylogenetic analyses of the *rep* genes (**Figure S1; Dataset S1H**). The nucleotide sequences of the *rep* genes and their corresponding amino acid sequences shared >95% identity within each pMIX plasmid group, except for pMI4 (**Table 1**). The clustering of the *rep* gene sequences belonging to the pMI4 plasmid group supported their inclusion in a single group, despite lower pairwise sequence identities (**Figure S1**). A total of 6 Rep protein families (RepA_N, Rep_3, Rep_1, Rep_trans, Rol_rep_N, and Rep_2) were encoded by the *rep* genes of the 9 pMIX groups (**Dataset S1H**). Specifically, the *rep* genes of pMI1 and pMI3 encoded protein families RepA_N and Rep_1, respectively. The *rep* genes of pMI2, pMI5, pMI6

and pMI7 encoded the protein family Rep_3. The *rep* genes of pMI4 and pMI8 encoded protein families Rep_trans and Rep_2, respectively. The *rep* of pMI9 was the only exception, encoding two Rep protein families, Rep_trans and Rol_rep_N (Table 1).

Each of the Dallas isolates harbored 1 to 5 of the pMIX plasmid groups, except the isolate 83 (ST17), which had none of them (Figure 1). Neither of the two reference strains harbored any of the pMIX plasmid groups. The pMI1 was the most frequently occurring pMIX plasmid group, identified in 75% of the isolates (Table 1). The pMI8 and pMI9 were the rarest ones, found in only isolates 51_1 (ST612) and 5 (ST18), respectively. The pMIX plasmid groups were not enriched in any specific ST. Plasmids within a very narrow size range (standard deviation 0.2 to 2.3) were grouped together (with the exception of pMI1), although plasmid size was not taken into consideration in assigning plasmid groups. The pMI1 consisted of plasmids size ranging from 23 to 62 kb (standard deviation 11.2).

To summarize, 9 new plasmid groups, pMI1 to pMI9, were defined for 110 plasmids based on *rep* gene sequences that encoded 6 different Rep protein families. Most of the Dallas isolates harbored 1 to 5 plasmids belonging to these newly defined plasmid groups.

Resistance genes were encoded in both chromosomes and plasmids. Different combinations of 12 resistance genes providing resistance against 5 classes of antibiotics (aminoglycosides, glycopeptides, macrolides, tetracycline, and trimethoprim) were identified among the Dallas isolates (Figure 3; Dataset S1; Supplemental text). The most commonly occurring resistance genes were *aac(6')/I-1*

(100% of isolates), *msr* (100%), *vanA* (95%), *tetM* (95%), *ermB* (85%), and *aph*(3') (85%). These resistance genes were encoded in the chromosome and mostly in two plasmid types, the pRUM-like and the megaplasmid. The pRUM-like plasmids predominantly encoded 4 of the resistance genes (*ant*(6)-*la*, *aph*(3')*III*, *vanA*, and *ermB*) conferring resistance against aminoglycosides, vancomycin and erythromycin (**Figure 3**). The *repUS7* plasmid (pMG1) was found among only 3 of the Dallas isolates (66 (ST18), 144-1 (ST18) and 51-4 (ST612)) and encoded *dfrG*, a less common resistance gene among the Dallas isolates, conferring resistance against trimethoprim.

All of the Dallas isolates, except for 163-1, encoded VanA-type vancomycin resistance genes, where *vanA* shared 99.9% (nucleotide) and 99.7% (amino acid) sequence identity with the reference sequence (GenBank AAA65956.1). The reference sequence is encoded within the transposon Tn1546 (GenBank M97297), originally described in the human fecal isolate *E. faecium* BM4147 collected from France in 1988³⁰ (**Dataset S1I**). A classical structure of the *vanA* gene cluster consisting of *vanHAX* and *vanYZ* encoded within Tn1546 elements was observed among the Dallas VREfm isolates, except for the isolate 17-1, which lacked the genes *vanY* and *vanZ*. The Tn1546 structure varied among the isolates in terms of insertion site (chromosome versus plasmid, IS elements, and point mutations, discussed further below.

The Tn1546 is borne on pRUM-like plasmids in the majority of isolates. The complete Tn1546 was carried by the pRUM-like plasmid in 42 of 46 *vanA*-positive isolates (**Figure 4**). The isolate 111 did not encode the entire Tn1546 in the pRUM-like plasmid; rather, it encoded *vanRS* in a 17 kb pRUM-like plasmid while *vanHAX*

was encoded within a separate 5 kb circular element (te111_5kb) (**Dataset S1F**).
Tn1546 was chromosomally encoded in 3 isolates, 53-1, 53-2 and 91, each belonging
to ST17, despite all 3 harboring a pRUM-like plasmid (**Table 2**).

Two toxin-antitoxin (TA) systems, *axe-txe* (TA_{axe-txe}) and a *relE/parE* family toxin
associated with an uncharacterized antitoxin component (TA_{relE}), were observed
alone or simultaneously on the pRUM-like plasmids (**Table 2**). Both of the TA systems
were actively expressed in two representative VREfm isolates (**Figure S2**).

The 43 pRUM-like plasmids encoding Tn1546 or a fragment of Tn1546 (for isolate
111) were categorized into the following four groups based on the presence or
absence of replication (as categorized by PlasmidFinder) and stability modules on
their backbones (**Table 2, Figure 5**). A previous study utilized the term “pRUM
derivatives” where plasmids were characterized by analyzing *rep* and TA systems¹².

- i) Group1 (*rep*₁₇; *rep*₁₈; TA_{axe-txe}; TA_{relE}): The pRUM-like plasmids categorized
into group 1 harbored a mosaic replication module *rep*_{17/18} as well as both
of the stability modules TA_{axe-txe} and TA_{relE}. All the group 1 plasmids
encoded the *rep18b* (accession number CP018068), except isolate 113,
which encoded the *rep18a* (accession number AB158402). A total of 16 of
the 43 plasmids belonged to this group. The plasmids had a narrow size
range (44 to 48 kbp) and were carried by VREfm isolates from 4 genetic
backgrounds (ST17, ST18, ST1383 and ST80).
- ii) Group 2 (*rep*₁₇; TA_{axe-txe}; TA_{relE}): A total of 17 of 43 plasmids harbored the
replication module *rep*₁₇ and both of the stability modules TA_{axe-txe} and
TA_{relE}, and were categorized into group 2. The group 2 plasmids could be

classified further based on size and genetic background of the isolates carrying them. Group 2a plasmids (29-45 kb) were carried by isolates belonging to ST18 (5, 66, 97-1, 124-1, 124-2, 124-3, and 162), ST80 (16-1, 16-2), ST17 (93-2), and ST612 (51-4), whereas larger (57-63 kb) group 2b plasmids were present in all 6 isolates belonging to ST664 (51-1, 87-1, 93-1, 93-2, 155, and 160). The reference strain ATCC® 700221™ harbored the group 2a pRUM-like plasmid. Sequence analysis revealed that Group 2b plasmids possess a set of genes that are specific to this group and present in no other group (**Dataset S1J**), including a ~7.5 kb region encoding helix-turn-helix XRE-family like proteins (cd00093) inserted immediately upstream of the *van* gene cluster (**Figure S3**).

iii) Group 3 (rep_{17} ; TA_{relE}): The group 3 plasmids possessed the replication module rep_{17} and the stability module TA_{relE} . A total of 5 of 43 plasmids were clustered into this group with the size range of 36-42 kb. They were carried by ST736 isolates (isolates 9-1, 9-2, 52, and 55) and one ST1703 isolate (isolate 121).

iv) Group 4 (rep_{17}): Group 4 plasmids encoded the replication module rep_{17} and no known stability modules ($TA_{N/A}$). A total of 5 of 43 plasmids with sizes ranging from 17 to 35 kb belonged to this group. These plasmids were carried by isolates belonging to the novel genetic background ST1703 (111, 137, 154-1, and 158) and by a single ST412 isolate (17-1).

It is interesting to note that the isolate 163-1, which belongs to CC17 but lacks vancomycin resistance genes, harbored a 25 kb pRUM-like plasmid. This pRUM-like plasmid did not harbor any part of the Tn1546 and did not belong to any of the plasmid sub-groups defined above. The presence of the modules rep_{17} and $TA_{axe-txe}$

and the absence of both of the modules *rep18* and TA_{relE} was unique among the Dallas isolates but was described before in the originally described pRUM plasmid in 2003³¹.

The robustness of the *rep*/TA typing system was confirmed using two approaches: a core genome phylogeny based on sequence variation in 7 pRUM-like core genes (*rep17*, IS1216, *vanS*, ImpB/MucB/SamB family protein, hypothetical protein, plasmid partitioning protein, and replication control protein PrgN) (**Figure S5**), and an analysis of gene presence/absence across 40 pRUM-like plasmids using Roary¹⁹ (**Dataset S1J**). The Group 2a pRUM-like plasmid from the reference strain *E. faecium* ATCC® 700221™ was included in the analysis; while 4 of the pRUM-like plasmid sequences (71-2, 100-2, 124-2, and 124-3) were not included due to sequence redundancy. The pRUM-like plasmids were clearly divided into two branches in the phylogenetic tree (**Fig S8**). Interestingly, these two groups corresponded to the presence/absence of the TA_{axe-txe}. Members of the plasmid groups 1, 2a and 2b clustered together, while those of group 3 and 4 were placed into a separate branch. The gene presence/absence heatmap (**Dataset S1J**) of 40 pRUM-like plasmids clearly identified genes specific to certain plasmid groups, other than the two stability and replication modules.

Dallas VREfm harbored 17 variants of Tn1546. The Tn1546 region of the Dallas isolates varied from the typical structure consisting of ORF1 (transposase), ORF2 (resolvase) and *van* genes flanked by IR_L (inverted repeat, left) and IR_R (inverted repeat, right) (**Figure 6**). A total of 17 different sequence variations in the Tn1546 were observed among 45 of the Dallas isolates (**Figure 6, Tables 3-4**). These 17 variations were classified using a nomenclature system developed previously¹³ based

on the presence of IS elements, point mutations, and deletions relative to the reference Tn1546. VREfm belonging to the same Tn1546 group shared 99.8%-100% nucleotide sequence identity in the Tn1546 region. The isolate 111 was excluded from this grouping since it harbored a fragmented Tn1546 in a pRUM-like plasmid and a circular entity (**Dataset S1F, Figure S2**). The Tn1546 variation was mostly associated with the distribution of 7 different IS elements (IS1251, IS1216, ISEfa16, ISEfa5, ISEfa17, IS256, ISEfa18), three of which are novel to the Dallas isolates. Most of the Tn1546 groups described in this study are novel, except for group BC2, which was reported previously in Poland¹³. The most predominant Tn1546 groups were BC9 (n=14) and J (n=6). The BC9 variety was observed in ST17 (n=9) and ST18 (n=5) isolates. The reference strain ATCC® 700221™ harbored the C2 structural variant of Tn1546 that was described previously¹³ but is not present in our Dallas VREfm isolates.

Identical Tn1546 groups were observed for isolates collected from the same patient (for example, all three VREfm isolates from patient 124 harbored Tn1546 type BCCL), except patients 51 and 93 (**Table 3**). Isolates 51-1 and 51-4 harbored Tn1546 type J and BC2 on the plasmid backbones 2b and 2a respectively. Isolates 93-1 and 93-2 harbored the same Tn1546 variety J on pRUM-like plasmid group 2b while 93-3 harbored BCM on plasmid group 2a. Interestingly the 51-1 (ST664) and 51-4 (ST612) isolates belonged to different clades in the phylogenetic tree. The isolates 93-1 (ST664) and 93-2 (ST664) clustered together but 93-3 (ST17) was placed in a separate clade. It is possible that pRUM-like vancomycin resistance plasmids were acquired by different *E. faecium* genotypes in two separate events in each of these patients.

Specific structural variations of the Dallas Tn1546 elements are described below.

The 3' but not 5' boundary of the Tn1546 was conserved among the Dallas isolates. Each of the 17 Tn1546 groups preserved the IR_R boundary of Tn1546 but lacked the IR_L boundary (**Figure 6**). The only exception was the isolate 17-1 (group BB CD) that lost both of the IR boundaries due to insertion of IS1216. The complete or partial deletion of ORF1 (transposase) and ORF2 (resolvase) was observed in all of the Tn1546 groups. This might decrease the likelihood of these Tn1546 to disseminate to other isolates without a plasmid as a vehicle.

IS1251 and IS1216 were the most common IS elements. IS1251 was present in 84.4% (38 of 45) of the isolates (**Table 3**). It was inserted at Tn1546 position 5820 in the *vanS-vanH* intergenic region, in reverse orientation to the *van* genes (**Figure 6**). The IS1251 generated an 8 bp direct repeat (5'-AAAATTAT-3') in the regions where it was inserted. IS1216 was inserted at either the 5' end of Tn1546 in 71.1% (32 of 45) or in the intergenic region of *vanX-vanY* in 17.8% (8 of 45) of the isolates.

The *van* genes and *vanR* and *vanH* promoters were intact in most isolates, despite IS element insertions. IS1251 insertion at Tn1546 position 5820 did not disrupt the two 12 bp VanR binding sites located at positions 5846 and 5879 (**Figure S4**). The 70 bp VanR-interacting region (-110 to -41) from positions 5935 to 6004 is also intact. The 197 bp long proposed *vanR* promoter region (-173 to +24) is located at the position 3,803 to 3,999 bp of the Tn1546. The isolate 137 (Tn1546 group BCK) has an ISEfa16 inserted at the position 3664 to 3763 that was responsible for a 99 bp deletion of resolvase at 3' region but the *vanR* promoter region remained uninterrupted. The *vanR* promoter region including the 12 bp consensus VanR

binding site is disrupted in the isolates 51-1, 87-1, 93-1, 93-2, 155, 160 from group J because of the insertion of IS1216 at position 3914. The *vanR* promoter region for other isolates remained uninterrupted.

Deletion of *vanR* and *vanS* was observed in the isolate 91 (variant group D1). Expression of the *vanHAX* genes may be constitutive instead of inducible in this isolate. No mutations were identified in *vanR* and *vanS* in the rest of the groups.

Two point mutations in *vanHAX* were fixed in the population. All isolates harbored a point mutation (T7658C) in *vanA* that generated a valine to alanine substitution (**Table 3**) and a second point mutation (G8234T) in *vanX* that generated a lysine to asparagine substitution. 62.2% (28 of 45) of the isolates harbored a third point mutation (C9692T) in *vanY* that was also present in the reference strain ATCC® 700221™. This point mutation generates a proline to leucine substitution in VanY. Isolates 53-1 and 53-2 (group BC10) had point mutations C6513A and G6514A in *vanH* that generated an arginine to lysine substitution. Isolate 144-1 harbored a unique G9566A point mutation in *vanY* that conferred a glycine to glutamic acid substitution. No mutations in *vanZ* were observed in any isolates.

Discussion

This study provided a snapshot of fecal VREfm genotypes among hospitalized patients in Dallas, Texas. To our knowledge, this was the first whole-genome epidemiological analysis of fecal VREfm isolates from patients without an active infection due to VREfm in this area.

Genetic diversity was observed among all of the 47 isolates analyzed. The most prevalent VREfm isolates, comprising 50% of the population, belonged to ST17 and ST18. These STs were previously reported to be responsible for outbreaks in countries including Portugal¹², Spain³², Ireland³³, Denmark³⁴, Columbia³⁵, Brazil³⁶, Canada³⁷. The other 8 STs, identified among our collection (ST1383, ST612, ST412, ST80, ST1703 and ST664), were single, double, or triple locus variants of ST17. The ST1703 was a novel sequence type described in this study. Notably, ST182, which caused an outbreak in San Antonio, Texas in the 1990s³⁸, was not detected in our collection. The isolate 163-1 was identified as a presumptive VREfm from clinical collection and was co-isolated with a *vanA*-positive VREfm not sequenced in this study¹⁶; but was sensitive to vancomycin by broth microdilution and negative for *vanA* and *vanB* by PCR screening¹⁶. This isolate was incorporated in our study despite being sensitive to vancomycin because it provided a reference point to compare with other Dallas isolates. This isolate generated its own clade, separate from the *vanA*-positive isolates in the phylogenetic tree, and its genetic background, ST262, has not been reported as VREfm to our knowledge. However, interestingly, the ST262 is a triple locus variant of ST17 (*pstS*-7, *atpA*-7, and *gyd*-5) and still belongs to the CC17. To conclude, VREfm isolates belonging to a combination of previously described and new genetic backgrounds colonize Dallas, Texas, patients.

The number of core genes identified among the Dallas isolates was 1706, which is much higher than the previously reported 1344⁵ and 955³⁹ core genes described by other groups. The lower number of samples (47 isolates) collected from a specific source (hospitalized, high-risk patients with fecal VREfm colonization from Dallas, Texas) is likely responsible for the larger number of core genes in our current study. The previous studies utilized larger sample sizes (73⁵ and 1,644³⁹), were collected

from different sources including hospitalized patients, healthy human volunteers, and animals, and included both VREfm and non-VREfm.

Completely closed plasmid sequences were achieved in this study by producing high quality genome assemblies. This made it possible to perform comparative analyses of entire plasmid sequences, which is a significant advantage compared to many previously employed techniques to analyze resistance plasmid diversity, including PCR-based typing schemes, restriction fragment length polymorphism (RFLP) comparisons, and hybridization¹².

All the plasmid-borne Tn1546 in Dallas isolates were encoded in pRUM-like plasmids that were clustered into 4 subgroups (1, 2a, 2b, 3, and 4), defined by the presence of the replication module *rep17* and a combination of other replication and stability modules. Speculatively, deletion or insertion of TA_{axe-txe} drove the evolution of the group 3 and 4 pRUM-like plasmids. The vancomycin-sensitive isolate 163-1 also harbored a pRUM-like plasmid, without Tn1546, that did not belong to any of the subgroups described in this study. This striking homogeneity in the plasmid backbone encoding *vanA* was surprising, because previous studies identified a mixture of different plasmids (pRUM-like, Inc18, mega-plasmid pLG1) as being responsible for vancomycin resistance in clinical isolates collected from a certain geographic region¹². However, our study differs from these previous studies in temporal and geographic features of sample collection. In previous studies, samples were collected over 2 to 12 years^{12,13,40} and from different regions of countries including Portugal¹², Poland¹³ and Germany⁴⁰.

The patients from whom the isolates in this study were collected resided in different areas of Dallas and in unconnected facilities including nursing homes and personal residences. The similarities among them were that they were high-risk patients, having recently been in a healthcare environment or having antibiotic exposure. Therefore, the presence of only one type of plasmid backbone carrying *vanA* in the Dallas isolates is striking. However, the existence of a common source of VRE acting as a reservoir and transferring a vancomycin resistance plasmid to *E. faecium* of different genetic background is an unlikely scenario since the pRUM backbone sequences and Tn1546 structure varied.

The phylogenetic analysis of pRUM-like plasmids in this study suggested an evolutionary event driven by either insertion or deletion of the TA_{axe-txe} which was originally described in the pRUM plasmid³¹. pRUM-like plasmids lacking TA_{axe-txe} were reported previously¹² although the possible implication of absence of TA_{axe-txe} on pRUM-like plasmid sub-grouping was not taken into consideration. The current study suggests a possibility of evolution of the pRUM-like plasmids by deletion of the TA system *axe-txe* since they have a second active TA_{relE}. pRUM-like plasmid groups 3 and 4, defined in this study, may be descendants of groups 1 and 2 in Dallas-area VREfm (**Figure S5**). Whether this impacts the maintenance of these plasmids in the absence of antibiotic selection, or their transfer efficiencies, requires further experimental study.

A total of 17 variations of the Tn1546 region, defined by the presence of different IS elements, point mutations, and deletions, were observed among the Dallas isolates. In absence of a strict guideline for the nomenclature of the Tn1546 variants, a recent one described in 2017¹³ was followed in this study. A slight variation in Tn1546

nomenclature was utilized in 2016 where variants with *IS1251* were designated as “type F” or “US hospital type”⁹, which are designated as group “C” in the current study. Three newly identified IS elements (*ISEfa16*, *ISEfa17*, and *ISEfa18*) within the *Tn1546* were observed in this study. Moreover, two point mutations within *Tn1546*, T7658C and G8234T, which were ubiquitous among the Dallas VREfm isolates, were reported previously but never in the same *Tn1546* group¹³. The point mutation T7658C was reported in the *Tn1546* groups BC5, BC1, BC2, BC3, and BC4 while G8234T was reported in groups A2, C1, C2, BB1, BB2, BC1, BC2, BC3, BC4, BC5, BBB1, and BBB2¹³.

The *Tn1546*-borne vancomycin resistance genes were carried by pRUM-like plasmids, which was categorized into plasmid groups 1, 2a, 2b, 3, and 4 based on replication and stability modules. Both of the pRUM-like plasmid groups 1 and 2a harbored the *Tn1546* variation BC9 (**Figure 7**). This was the only example of two different plasmid groups carrying the same *Tn1546* variation. Therefore, the plasmid backbone may have diverged into groups 1 and 2a after acquisition of the BC9 type *Tn1546*. The same variation of *Tn1546*-type transposon was never observed for the plasmid groups 2b, 3 and 4. Therefore, an independent evolution of *Tn1546* and the plasmid backbone can be suggested for the rest of the resistance plasmid groups.

The resistance plasmid group 1 harboring the same *Tn1546* variant, BC9, in two different genetic backgrounds, ST17 (isolate 1, 20, 57, 83, 103, 107-1, 131-2, 148, and 161) and ST18 (isolate 100-1 and 100-2), was observed. Horizontal transfer of this specific plasmid between ST17 and ST18 is possible (**Figure 7**). Similarly, horizontal transfer of plasmid group 1 with *Tn1546* variant BC11 between isolates belonging to ST18 and ST1383 may have occurred, as well as horizontal transfer of

plasmid group 2a with Tn1546 variant BC2 between isolates belonging to ST80 and ST612 (**Figure 7**). No other evidence of horizontal transfer of pRUM-like plasmids among different VREfm genetic backgrounds was observed. Different plasmid groups present in isolates with the same genetic background (group 1, 2a in ST17, ST18, ST80; group 3, 4 in ST1703) suggested independent acquisition events of the resistance plasmids within those STs. The resistance plasmid groups 1 and 2a were observed in the common STs 17 and 18. The other resistance plasmid groups 2b, 3 and 4 were carried by the less prevalent STs. So, a combination of both horizontal transfer and independent acquisition events followed by IS element-driven Tn1546 evolution shaped the VanA-type VREfm population in this area.

To conclude, the molecular diversity of multidrug resistant VREfm isolates in Dallas, Texas, collected in the year 2015 was documented. VREfm with a novel genetic background, ST1703, was reported in this study. No clonal spread of any particular VREfm genotype could be detected in this isolate collection. However, strikingly, vancomycin resistance genes were borne by pRUM-like plasmids in almost all isolates despite the diversity of patient characteristics. Diversity of Tn1546 was described and new variations involving 3 novel IS elements (ISEfa16, ISEfa17 and ISEfa18) were also described among the VREfm collection. Horizontal transfer of resistance plasmids between ST17-ST18 and ST80-ST612 was speculated since VREfm of these genetic backgrounds carried the same variation of Tn1546 on the same plasmid backbone. Independent acquisition of resistance was observed in the rest of the isolates. Therefore, it was concluded that VREfm representatives from different sources carrying unique Tn1546 generated the demography of VREfm in Dallas, Texas.

Acknowledgement

This project was supported by R01AI116610 and the Cecil H. and Ida Green Chair in Systems Biology Science to K.P. We gratefully acknowledge Dr. Joslyn Pribble and the Microbiology lab at Methodist Health System, without whom this work would not be possible.

References

1. Bonten MJM, Willems R, & Weinstein RA. Vancomycin-resistant enterococci: why are they here, and where do they come from? *Lancet Infect. Dis.* **1**, 314–325 (2001).
2. Uttley AH, Collins CH, Naidoo J, & George RC. Vancomycin-resistant enterococci. *Lancet* **9**, 8575–6 (1988).
3. Sahm DF, Kissinger J, Gilmore MS, Murray PR, Mulder R, Solliday J, & Clarke B. In vitro susceptibility studies of vancomycin-resistant *Enterococcus faecalis*. *Antimicrob. Agents Chemother.* **33**, 1588 LP – 1591 (1989).
4. Arias CA & Murray BE. The rise of the Enterococcus : beyond vancomycin resistance. *Nat. Publ. Gr.* **10**, 266–278 (2012).
5. Lebreton F, Van Schaik W & Manson A. Emergence of Epidemic Multidrug-Resistant *Enterococcus faecium*. *Am. Soc. Microbiol.* **4**, 1–10 (2013).
6. Palmer KL, Godfrey P, Griggs A, Kos VN, Zucker J, Desjardins C, Cerqueira G, Gevers D, Walker S, Wortman J, Feldgarden M, Haas B, Birren B, & Gilmore MS. Comparative Genomics of Enterococci: Variation in *Enterococcus faecalis*, Clade Structure in *E. faecium*, and Defining Characteristics of *E. gallinarum* and *E. casseliflavus*. *MBio* **3**, 1–11 (2012).
7. Homan WL, Tribe D, Poznanski S, Li M, Hogg G, Spalburg E, Van Embden JD, & Willems RJ. Multilocus sequence typing scheme for *Enterococcus faecium*.

- 687 *J. Clin. Microbiol.* **40**, 1963–1971 (2002).
- 688 8. Lee T, Pang S, Abraham S, & Coombs GW. Antimicrobial-resistant CC17
689 *Enterococcus faecium*: The past, the present and the future. *J. Glob.* **16** 36–47
690 (2019) .
- 691 9. Freitas AR, Tedim AP, Francia MV, Jensen LB, Novais C, Peixe L, Sánchez-
692 Valenzuela A, Sundsfjord A, Hegstad K, Werner G, Sadowy E, Hammerum
693 AM, Garcia-Migura L, Willems RJ, Baquero F, Coque TM. Multilevel population
694 genetic analysis of vanA and vanB *Enterococcus faecium* causing nosocomial
695 outbreaks in 27 countries (1986–2012). *J. Antimicrob. Chemother.* **71**, 3351–
696 3366 (2016).
- 697 10. Partridge SR, Kwong SM, Firth N, Jensen SO. Mobile Genetic Elements
698 Associated with Antimicrobial Resistance. *Clin. Microbiol. Rev.* **31**, (2018).
- 699 11. Courvalin P. Vancomycin resistance in gram-positive cocci. *Clin. Infect. Dis.*
700 *an Off. Publ. Infect. Dis. Soc. Am.* **42 Suppl 1**, S25–34 (2006).
- 701 12. Freitas A, Novais C, Tedim AP, Francia MV, Baquero F, Peixe L, Coque TM.
702 Microevolutionary events involving narrow host plasmids influences local
703 fixation of vancomycin-resistance in *Enterococcus* population. *PLoS One* **8**,
704 e60589 (2013).
- 705 13. Wardal E, Kuch A, Gawryszewska I, Żabicka D, Hryniewicz W, Sadowy E.
706 Diversity of plasmids and Tn1546-type transposons among VanA
707 *Enterococcus faecium* in Poland. *Eur. J. Clin. Microbiol. Infect. Dis.* **36**, 313–
708 328 (2017).
- 709 14. Loman NJ, Quick, J, & Simpson JT. A complete bacterial genome assembled
710 de novo using only nanopore sequencing data. *Nat. Methods* **12**, 733–735
711 (2015).
- 712 15. Bayliss SC, Hunt VL, Yokoyama M, Thorpe HA, & Feil EJ. The use of Oxford

713 Nanopore native barcoding for complete genome assembly. *Gigascience* **6**, 1–
714 6 (2017).

715 16. Ping S, Mayorga-Reyes N, Price VJ, Onuoha M, Bharadwaj P, Rodrigues M,
716 Owen J, Araya DP, Akins RL, Palmer KL. Characterization of presumptive
717 vancomycin-resistant enterococci recovered during infection control
718 surveillance in Dallas, Texas. *medRxiv* 2020.10.09.20209569 (2020).
719 doi:10.1101/2020.10.09.20209569

720 17. Dutka-Malen S, Evers S, & Courvalin P. Detection of glycopeptide resistance
721 genotypes and identification to the species level of clinically relevant
722 enterococci by PCR. *J. Clin. Microbiol.* **33**, 24–27 (1995).

723 18. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**,
724 2068–9 (2014).

725 19. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, Fookes M,
726 Falush D, Keane JA, Parkhill J. Roary: rapid large-scale prokaryote pan
727 genome analysis. *Bioinformatics* **31**, 3691–3 (2015).

728 20. Stamatakis A. RAxML Version 8: A tool for Phylogenetic Analysis and Post-
729 Analysis of Large Phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

730 21. Letunic I, & Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new
731 developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).

732 22. Jolley KA, Bray JE, & Maiden MCJ. Open-access bacterial population
733 genomics: BIGSdb software, the PubMLST.org website and their applications.
734 *Wellcome Open Res* **3**, doi: 10.12688/wellcomeopenres.14826.1. eCollection
735 (2018).

736 23. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L,
737 Møller Aarestrup F, Hasman H. PlasmidFinder and pMLST: in silico detection
738 and typing of plasmids. *Antimicrob. Agents Chemother.* **58**, 3895–903 (2014).

- 739 24. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web
740 server: 2018 update. *Nucleic Acids Res.* **46**, W200–W204 (2018).
- 741 25. Tamura K. & Nei M. Estimation of the number of nucleotide substitutions in the
742 control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol.*
743 *Evol.* **10**, 512–26 (1993).
- 744 26. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O,
745 Aarestrup FM, L. M. Identification of acquired antimicrobial resistance genes. *J*
746 *Antimicrob Chemother* **67**, 2640–2644 (2012).
- 747 27. Siguier P, Perochon J, Lestrade L, Mahillon J, & Chandler M. ISfinder: the
748 reference centre for bacterial insertion sequences. *Nucleic Acids Res.* **34**, D32-
749 6 (2006).
- 750 28. McKenney P, Ling L, Wang G, Mane S. & Pamer EG. Complete Genome
751 Sequence of *Enterococcus faecium* ATCC 700221. *Am. Soc. Microbiol.* **4**,
752 e00386-16 (2016).
- 753 29. ATCC. Multidrug-resistant and antimicrobial testing reference strains. (2019).
- 754 30. Arthur M, Molinas C, Depardieu F. & Courvalin P. Characterization of Tn1546,
755 a Tn3-related transposon conferring glycopeptide resistance by synthesis of
756 depsipeptide peptidoglycan precursors in *Enterococcus faecium* BM4147. *J.*
757 *Bacteriol.* **175**, 117–127 (1993).
- 758 31. Grady R. & Hayes F. Axe-Txe, a broad-spectrum proteic toxin-antitoxin system
759 apesified by multidrug-resistant, clinical isolate of *Enterococcus faecium*. *Mol*
760 *Microbiol* **47**, 1419–32 (2003).
- 761 32. Valdezate S, Miranda C, Navarro A, Freitas AR, Cabrera JJ, Carrasco G,
762 Coque TM, Jiménez-Romano E, Saéz-Nieto JA. Clonal outbreak of ST17
763 multidrug-resistant *Enterococcus faecium* harbouring an Inc18-like::Tn1546
764 plasmid in a haemo-oncology ward of a Spanish hospital. *J. Antimicrob.*

- 765 *Chemother.* **67**, 832–836 (2012).
- 766 33. Ryan L, O'Mahony E, Wrenn C, FitzGerald S, Fox U, Boyle B, Schaffer K,
767 Werner G, Klare I. Epidemiology and molecular typing of VRE
768 bloodstream isolates in an Irish tertiary care hospital. *J. Antimicrob. Chemother.*
769 **70**, 2718–2724 (2015).
- 770 34. de Been M, Pinholt M, Top J, Bletz S, Mellmann A, van Schaik W, Brouwer E,
771 Rogers M, Kraat Y, Bonten M, Corander J, Westh H, Harmsen D, Willems
772 RJ. Core Genome Multilocus Sequence Typing Scheme for High- Resolution
773 Typing of *Enterococcus faecium*. *J. Clin. Microbiol.* **53**, 3788–3797 (2015).
- 774 35. Rios R, Reyes J, Carvajal LP, Rincon S, Panesso D, Echeverri AM, Dinh A,
775 Kolokotronis SO, Narechania A, Tran TT, Munita JM, Murray BE, Planet PJ,
776 Arias C, Diaz L. Genomic Epidemiology of Vancomycin-Resistant
777 *Enterococcus faecium* (VREfm) in Latin America: Revisiting The Global VRE
778 Population Structure. *Sci. Rep.* **10**, 5636 (2020).
- 779 36. Camargo I LBC, Gilmore MS & Darini ALC. Multilocus sequence typing and
780 analysis of putative virulence factors in vancomycin-resistant and vancomycin-
781 sensitive *Enterococcus faecium* isolates from Brazil. *Clin. Microbiol. Infect.* **12**,
782 1123–1130 (2006).
- 783 37. McCracken M, Wong A, Mitchell R, Gravel D, Conly J, Embil J, Johnston L,
784 Matlow A, Ormiston D, Simor AE, Smith S, Hizon DR, Mulvey MR, on behalf of
785 the members of the Canadian Nosocomial Infection Surveillance Program.
786 Molecular epidemiology of vancomycin-resistant enterococcal bacteraemia:
787 results from the Canadian Nosocomial Infection Surveillance Program, 1999-
788 2009. *J. Antimicrob. Chemother.* **68**, 1505–1509 (2013).
- 789 38. Moreno F, Grota P, Crisp C, Magnon K, Melcher GP, Jorgensen JH, Patterson
790 JE. Clinical and molecular epidemiology of vancomycin-resistant *Enterococcus*

faecium during its emergence in a city in southern Texas. *Clin. Infect. Dis. an*
Off. Publ. Infect. Dis. Soc. Am. **21**, 1234–1237 (1995).

39. Arredondo-Alonso S, Top J, McNally A, Puranen S, Pesonen M, JPensar
J, Marttinen P, Braat JC, Rogers MRC, van Schaik W, Kaski S, Willems
RJL, Corander J, Schürch AC. Plasmids Shaped the Recent Emergence of the
Major Nosocomial Pathogen *Enterococcus faecium*. *MBio* **11**, (2020).
40. Bender JK, Kalmbach A, Fleige C, Klare C, Fuchs S, Werner G. Population
structure and acquisition of the vanB resistant determinant in German clinical
isolates of *Enterococcus faecium* ST192. *Sci. Rep.* **6**, (2016).

Table 1. Features of the pMIX group plasmids

Plasmid group	No. of isolates with plasmid group ¹	% of isolates harboring this plasmid ²	Plasmid size range (kb)	stdv of plasmid size within group	<i>rep</i> gene information				
					<i>rep</i> gene size (bp)	% identity within group (<i>rep</i> DNA) ³	% identity within group (<i>rep</i> aa) ⁴	<i>Rep</i> encoded Protein Family Id ⁵	pfam Accession ⁶
pMI1	35	74	23-62	11.2	1086	99.7	98.9	repA_N	PF06970.11
pMI2	32	68	6-8	0.3	951	100	100	Rep_3	PF01051.21
pMI3	17	36	3.6-4.3	0.2	990	100	100	Rep_1	PF01446.17
pMI4	11	23	4-11	2.3	972	79.4	67.9	Rep_trans	PF02486.19
pMI5	5	11	83-106	8.7	804	95.40	95.9	Rep_3	PF01051.21
pMI6	4	9	4	0	924	100	100	Rep_3	PF01051.21
pMI7	4	9	3	0	720	100	100	Rep_3	PF01051.21
pMI8	1	2	2	N/A	603	N/A	N/A	Rep_2	PF01719.17
pMI9	1	2	70	N/A	1215	N/A	N/A	RoI_Rep_N	PF18106.1
								Rep_trans	PF02486.19

¹Specifies the total number of isolates among the 47 Dallas isolates harboring the particular pMIX group plasmids.

²The % of isolates harboring the specific plasmid group.

^{3,4}The “% identity (DNA)” and the “% identity (aa)” specifies the % identical sites calculated from the MUSCLE sequence alignment of DNA and amino acid of the *rep* genes belonging to the same pMIX group, respectively.

^{5,6} The *rep* gene encoded protein family (pfam) and the corresponding pfam accession numbers, respectively.

836

Table 2. Features of the pRUM-like plasmid groups

plasmid groups	MLST ¹	Isolate	<i>rep</i> ²		TA ³		size (bp)	Size (kbp) (Avg ±stdev) ⁴
			rep17	rep18	axe-txe	relE		
Group 1	ST17	1					44,323	44 ± 1.9
		20					44,443	
		43-1					44,730	
		57					44,421	
		83					44,442	
		103					48,758	
		107-1					41,004	
		131-2					43,860	
		148					45,965	
		161					44,443	
	ST18	100-1					43,236	
		100-2					43,236	
		144-1					47,773	
	ST1383	71-1					44,420	
		71-2					44,420	
	ST80	113		a			41,153	
Group 2a	ST18	5					45,647	37 ± 4.3
		66					38,373	
		97-1					39,713	
		124-1					37,877	
		124-2					37,877	
		124-3					37,877	
		162					34,583	
	ST80	16-1					41,600	
		16-2					33,441	
	ST17	93-3					34,144	
	ST612	51-4					29,409	
Group 2b	ST664	51-1					61,762	59 ± 3
		87-1					57,116	
		93-1					63,108	
		93-2					63,108	
		155					57,134	
		160					57,135	
Group 3	ST736	9-1					38,857	38 ± 2.6
		9-2					38,857	
		52					36,284	
		55					36,284	
	ST1703	121					42,671	
Group 4	ST1703	111					17,982	28 ± 7.4
		137					35,487	
		154-1					22,729	
		158					30,941	
	ST412	17-1					33,158	

837

838 ¹Multiple locus sequence type (MLST) of the isolates

^{2,3}The presence of *rep* and toxin-antitoxin (TA) genes are indicated by blue boxes.

The pRUM-like plasmid from isolate 113 harbored *rep18a* (reference gene Accession no. AB158402) and is symbolized as “a” in the corresponding box. The *rep18* for rest of the resistance plasmids were *rep18b* (reference gene Accession no. CP018068).

⁴The average size of the plasmids for each plasmid group type is shown along with the standard deviation (Avg. $\pm$ stdev).

Table 3. Features of the Tn1546 groups observed among the Dallas isolates

pRUM-like plasmid group	Tn1546 Group	Isolate	Identity within Tn1546 group ¹	No. of isolates in Tn1546 group	IS element ²	Point mutations ³				
						vanH (6513, 6514 CG--> AA) Arg --> Lys	vanA (7658 T-->C) Val-->Ala	vanX (8234 G-->T) Lys-->Asn	vanY (9692 C-->T) Pro-->Leu	vanY (9566 G-->A) Gly --> Glu
1	BCL	43-1	N/A	1	IS1216, IS1251, ISEfa17					
	BC8	113	N/A	1	IS1216, IS1251					
	BC11	71-1	99.80%	3	IS1216, IS1251					
		71-2								
		144-1								
		1								
	BC9	20	99.80%	14	IS1216, IS1251					
		57								
		83								
		100-1								
		100-2								
		103								
		107-1								
		131-2								
		148								
		161								
		5								
		66								
		162								
2a		BC2				16-1	100%	3	IS1216, IS1251	
	16-2									
	51-4									
	BBC	97-1	N/A	1	IS1216, IS1251					
	BCM	93-3	N/A	1	IS1216, IS1251, ISEfa18					
	BCCL	124-1	100%	3	IS1216, ISEfa17, IS1251, IS1251					
		124-2								
2b	J	124-3								
		51-1	100%	6	IS256					
		87-1								
		93-1								
		93-2								
		155								
160										
3	C3	52	100%	2	IS1251					
		55								
	BC6	9_1	100%	2	IS1251, IS1216					
		9_2								
4	BC7	121	N/A	1	IS1251, IS1216					
	BCB	154-1	100%	2	IS1216, IS1251, IS1216					
		158								
	BCK	137	N/A	1	ISEfa16, IS1251, IS1216					
	BBCD	17-1	N/A	1	IS1216, ISEfa5, IS1251, IS1216					
Chromosome integrated Tn1546	D1	91-1	N/A	1	ISEfa5					
	BC10	53-1	100%	2	IS1216, IS1251					
53-2										

¹ Nucleotide sequences of the Tn1546 region for the Tn1546 groups, including more than one isolate, were aligned and % identity was determined.

²The name of the IS elements present in the Tn1546 group are listed.

³Point mutations in the van gene cluster relative to the reference Tn1546 (accession number M97297). Presence of a point mutation is represented by a colored box.

875

876

877

878

879 **Figure 1. Plasmid content of Dallas *E. faecium* isolates.** The phylogenetic tree of
880 49 *E. faecium* strains is based on an alignment of SNPs in 1706 core genes. The
881 branches colored green, golden and red are supported by a bootstrap value of >70,
882 40-60, and <30 respectively. The reference genomes *E. faecium* 1,231,501 and
883 ATCC® 700221™ are labeled in red. Data for MLST, number of presumptive
884 plasmids (shown as ## for incomplete assemblies), and plasmid *rep* types are shown.
885 Plasmid replicon conserved domains i) RepA_N, ii) Rep_3, iii) Rep_trans, iv) Rep1, v)
886 Inc18, vi) Rep_2, vii) Rol_Rep_N (A), *rep* family (B), plasmid size ranges (C), and
887 reference plasmid names (D) present in each strain are shown. A dark purple box
888 indicates a chromosomally integrated plasmid whose size range was not determined.
889 The *rep17* and mosaic of *rep17/rep18* are indicated with the gray and light green
890 boxes, respectively. New plasmid groups identified in this study (pMI1 to pMI9) are
891 highlighted in light blue color.

892

893

Figure 2: Sequence type (ST) distribution of Dallas *E. faecium* isolates. A total of 10 STs were observed across 47 isolates. The frequency of STs (percent of total isolates) is shown.

Figure 3. Distribution of antibiotic resistance genes among Dallas isolates. Shown are 12 resistance genes (6'-N-aminoglycoside acetyltransferase, (*aac(6')**I-i*), bifunctional enzyme 6'-aminoglycoside acetyltransferase-2''-aminoglycoside phosphotransferase (*aac(6')**-aph(2'')*), aminoglycoside nucleotidyltransferase (*ant(6)-Ia*), *aph(3')**III*, *vanA*, *erm(B)*, *erm(T)*, *lnu*, acquired macrolide resistance-like protein (*msr*), *tet(M)*, *tet(L)* and dihydrofolate reductase (*dfrG*)), and the antibiotics against which they provide resistance to. A solid colored box indicates the presence of a certain resistance gene in the isolate. The boxes are color coded to indicate the location of the gene (chromosome or a certain type of plasmid).

Figure 4. Genetic map of a representative Group 2a pRUM-like plasmid. The 39 kb plasmid p97_1_39kb (accession number CP066571) from the Dallas VREfm isolate 97-1 is shown. Arrows represent genes. The genes are color-coded based on their predicted activity. Maintenance genes are predicted to be responsible for general maintenance of the plasmid within the bacterial host. Resistance genes provide antibiotic resistance.

Figure 5. Classification and distribution of pRUM-like plasmid groups. A) pRUM-like plasmids were classified based on the presence or absence (indicated by a + or -, respectively) of 2 replication modules (*rep17* and *rep18*) and 2 stability modules (toxin-antitoxin system; TA_{axe-txe} and TA_{relE}). **B)** Frequency of the pRUM-like plasmid groups among 43 Dallas isolates.

Figure 6: Tn1546 structural variations in the Dallas VREfm isolates. A reference Tn1546 structure that includes transposase, resolvase and *van* gene cluster flanked by *IR_L* and *IR_R* (inverted repeat left and right, respectively) is delineated at the top. The Tn1546 groups are represented below the reference Tn1546, and the name of the group (C3, BC2, BC6, BC7, BC8, BC9, BC10, BC11, BCB, BBC, BCK, BCL, BCM, BB CD, BCCL D1, J) is shown on the left. Red boxes represent deleted regions relative to the reference Tn1546. Vertical lines with numbers at the bottom represent the nucleotide position within the reference Tn1546. IS elements are represented by downward arrows. Black stars indicate point mutations. The figure is not drawn to scale.

Figure 7. Evolutionary relationships among pRUM-like plasmid groups, *Tn1546*, and VREfm genetic background. The resistance plasmid groups 1, 2a, 2b, 3, and 4 are represented on the left side. The *Tn1546* variations encoded in the corresponding plasmid backbone are displayed on the right side in white boxes. The numbers in parentheses represent total numbers of isolates harboring this *Tn1546* variant on this specific plasmid group belong to this genetic background (ST). The *Tn1546* variant BC9 in plasmid groups 1 and 2a is highlighted with pink color to indicate the unique event of the presence of same *Tn1546* variation in two different plasmid groups. The sequence type (ST) of VREfm isolates carrying the *Tn1546* variation on the specific plasmid backbone is indicated with blue ovals. A solid red arrow links the ST and the *Tn1546* variation carried by the corresponding plasmid backbone. The dotted black arrow indicates a possible horizontal transfer of resistance plasmid between VREfm isolates belonging to that genetic background.