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]#

[a]Department of Biological Sciences, University of Texas at Dallas, Richardson, Texas, USA

[b]Methodist 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 Tn*1546*-like elements on a pRUM-like plasmid backbone. New variants of the Tn*1546* were described that harbored a combination of 7 insertion sequences (IS) including 3 novel IS elements reported in this study (IS*Efa16*, IS*Efa17* and IS*Efa18*). The pRUM-like plasmids carrying the Tn*1546* 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 Tn*1546*-like elements and the *vanA* operon, which have evolved independently of the pRUM-like backbone.

53 **Importance**

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

66

67

68 **Introduction**

69 Originally described as a commensal of the healthy human gut, the enterococci
70 attracted attention in the 1990s due to their alarming multi-drug resistance[1].
71 Increasing cases of vancomycin-resistant enterococci (VRE), first reported in 1988 in
72 Europe[2] and then in USA in 1989[3] became a major concern for nosocomial infection[4].
73 In recent years, vancomycin-resistant *E. faecium* (VREfm) is considered to be the
74 most prominent VRE species, and hence is widely studied[4].

75

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

88

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

94 complex-17 (CC17) that was founded by the ST17[8]. The VREfm isolates involved in
95 hospital outbreaks in different parts of the world, within the time period 1986-2012,
96 represented 26 STs[9]. The most prominent VREfm isolates reported were ST17
97 (reported in Spain, China, Brazil), ST18 (reported in Canada, Poland, South America,
98 Turkey, Serbia) and ST78 (reported in Turkey, Paraguay, Belgrade, Italy). In USA,
99 VREfm of ST16 and ST17 were responsible for clonal outbreaks from 1990 to 2005.
100 Generally, a single ST of VREfm is associated with an outbreak in a certain
101 geographic region, while globally, outbreaks are caused by multiple different STs. For
102 example, ST25 (France, 1987), ST114 (Brazil, 1998), ST18 (UK, 1992) and ST16
103 (Netherland, 2002) were reported outbreak strains from different parts of the world[9].
104 Epidemiological differences in diverse geographical areas[9] reiterate the importance of
105 tracking VREfm genotypes.
106
107 Vancomycin resistance genes are encoded within mobile elements in VREfm, and
108 can be horizontally disseminated[10]. The most commonly occurring vancomycin
109 resistance type worldwide, currently, is VanA-type resistance, which is typically
110 encoded within the element Tn*1546*[11]. Broadly speaking, the Tn*1546* may be
111 chromosomally integrated or plasmid-borne. Different plasmid families carrying
112 Tn*1546* including Inc18, pRUM, and pLG1 have been reported worldwide. Hybrids of
113 pRUM-Inc18, Inc18-pHTβ, pRUM-pLG1 family plasmids carrying *vanA* have been
114 reported as well[9]. Structural variations of Tn*1546* due to the presence of insertion
115 sequences (IS) like IS*1251* (USA, Australia, Europe), IS*1216* (Europe), IS*1485* (Saudi
116 Arabia) and IS*Ef1* (Germany, Portugal, Poland) have also been reported. Previous
117 studies determined the Tn*1546* structural variation by PCR mapping and then
118 sequencing of the overlapping fragments[9],[12],[13]. However, in the absence of a
119 completely closed genome assembly, conclusive links between *vanA*, Tn*1546*

120 variants, and specific plasmid backbones in VREfm is difficult to achieve. Surveillance
121 of resistance elements is important because changes in host range and transfer
122 frequency can evolve, which could impact clinical care. Utilization of whole genome
123 sequencing for epidemiologic study provides an opportunity for multilevel molecular
124 analysis that includes detection of novel mobile genetic elements (MGE) and
125 monitoring changes in resistance gene vectors. The advent of long-read next
126 generation sequencing (NGS) by Oxford Nanopore Technology (ONT) made it
127 possible to resolve long repeat regions and generate genome assemblies of a large
128 number of bacterial isolates[14] in a cost effective manner. The combination of long
129 (ONT) and short (Illumina) reads generates high quality completely closed genome
130 assemblies that gives access to fully assembled mobile genetic elements[15].
131
132 Much of the focus of VREfm genomics has been on infection-derived isolates. There
133 is a comparative lack of information regarding genotypic diversity of fecal (i.e., non-
134 infection-derived) VREfm isolates that colonize hospitalized patients in a given
135 population. These represent reservoirs from which future hospital outbreaks could
136 emerge. The current study aims to address this knowledge gap by analyzing fecal
137 VREfm isolates collected from hospitalized patients in Dallas, Texas, using a
138 combination of ONT and Illumina sequencing.
139
140 **Materials and Methods**
141 **Bacterial strains analyzed in this study**. The isolates used in this study (**Dataset**
142 **S1A**) were cultured from surveillance rectal swabs of high-risk patients on hospital
143 admission within a span of two months (August 2015 to October 2015), and their
144 clinical collection and initial characterization was previously reported[16]. Briefly, rectal
145 swabs from patients admitted with at least one of the following risk factors

146 (hospitalization for ≥2 consecutive nights in the preceding 30 days, transferred from
147 another medical facility, residence in a nursing home or extended/long term care
148 facility, or the presence of decubitus ulcer or a draining wound) were cultured on
149 Spectra VRE (Remel) agar as a part of routine infection control surveillance. Instead
150 of disposal after clinical procedures were complete, cultures positive for presumptive
151 VRE growth were coded numerically, de-identified, and transferred to the University of
152 Texas at Dallas for further analysis. Isolates were coded by de-identified patient
153 number. If multiple colonies were collected from a Spectra VRE plate, this was
154 indicated by hyphenated numerics. For example, the VRE isolates 124-1, 124-2 and
155 124-3 were collected from the same patient (de-identified patient #124). In our
156 previous study of these isolates, presumptive VREfm were confirmed for species by
157 *ddl* typing and for the presence of *vanA* or *vanB* genes by PCR[16,17]. All isolates used
158 in this study, with the exception of 163-1, grew on brain heart infusion (BHI) agar
159 supplemented with 256 μg/mL vancomycin. The vancomycin MIC of isolate 163-1
160 was previously determined to be 2 μg/mL by broth microdilution[16].
161
162 **Genome sequencing and assembly.** Methods for DNA isolation, MinION
163 sequencing, Illumina sequencing, and hybrid genome assembly are described in the
164 Supplemental Text.
165
166 **Phylogenetic analyses and MLST**. Genome assemblies were annotated with
167 PROKKA (v 1.12)[18]. A sequence alignment was generated for 1706 core genes by
168 ROARY (v 3.12)[19] using PRANK where the paralogs were not split. A phylogenetic
169 tree was built by RAxML (v 8.2.12)[20] with rapid bootstrapping (1000 inferences) using
170 the GTR+GAMMA model and subsequent ML search on the core gene alignment

171 where the outgroup was specified to be the reference genome *E. faecium* 1,231,501
172 (GenBank accession number ACAY01). The phylogenetic tree was visualized with
173 iTOL (v4)[21]. MLST was determined by the *E. faecium* MLST database
174 (https://pubmlst.org/efaecium/)[22,7]. The novel sequence type ST1703 and novel allele
175 number for *gyd* (allele-70) was assigned for 5 VREfm isolates (111, 121, 137, 154-1,
176 158) upon request from the database.
177
178 **Plasmid detection and typing**. Each contig of <2 Mbp in size in a closed genome
179 assembly was considered a putative plasmid. *In silico* detection and typing of
180 plasmids was performed with PlasmidFinder[23] (v 2.1). Putative plasmids in closed
181 genome assemblies that were not classifiable by PlasmidFinder were analyzed further
182 (see below).
183
184 **Definition of pMIX plasmid groups**. Putative plasmid sequences that were
185 unclassifiable by PlasmidFinder were analyzed as follows. Presumptive *rep* genes
186 were identified from plasmid annotations and analyzed by NCBI Conserved Domain
187 Analysis (CDS). Confirmed *rep* genes encoded protein domains that matched known
188 Rep protein families with e-values of e^{-10} to e^{-78}. The Rep sequences were also
189 scanned by HMMER (v 2.41.1)[24] to analyze pfam domains. Multiple sequence
190 alignment of the nucleotide and predicted amino acid sequences of *rep* genes were
191 performed with MUSCLE (v 3.8.424) where the "group sequences by similarity" option
192 was selected. A neighbor-joining tree was built based on the sequence alignment
193 using the "Tamura Nei" genetic distance model[25]. The tree was visually scrutinized in
194 order to sort the plasmids into 9 groups harboring *rep* genes with at least 95% identity
195 in both nucleotide and predicted amino acid sequence (except for pMI4; see Results).

196 Each unique *rep* group of plasmids was given a name of format pMIX, where the "X"
197 is a numeric number from 1 to 9.
198
199 Eight closed circular DNA elements of varying sizes (3 to 72 kb) could not be
200 classified by PlasmidFinder nor by the pMIX *rep* typing scheme described above.
201 Analysis of these elements is described in the Supplemental Text.
202
203 **Transcriptional activity of predicted toxin-antitoxin (TA) systems**. Transcriptional
204 activity of the TA systems, $TA_{axe\text{-}txe}$ and TA_{relE}, was determined by RT-qPCR for the
205 VREfm isolates 1 and 5, described in the Supplemental Text. The primer sequences
206 used for qPCR are provided in **Dataset S1B.**
207
208 **Resistance gene identification**. Acquired antimicrobial resistance genes were
209 detected with ResFinder 3.1[26].
210
211 **Tn*1546* analysis.** The Tn*1546* nucleotide sequence described in *E. faecium* BM4147
212 (accession number M97297) was used as a reference to identify variations occurring
213 in the Tn*1546* identified in this study. The Tn*1546* were classified using a previously
214 described nomenclature[13]. The presence of insertion sequence elements (IS
215 elements) was denoted with a one-letter code, shown in parentheses: IS*1216* (B),
216 IS*1251* (C), IS*Efa5* (D), IS*256* (J), IS*Efa16* (K), IS*Efa17* (L), IS*Efa18* (M). The
217 combination of IS elements within the transposon is described by a two or three letter
218 code; i.e., group 'BC' possesses one each of IS*1216* and IS*1251*. Arabic numerals,
219 following the alphabet code, indicate differences due to point mutations or different IS
220 element insertion sites. For the group 'BC', a previous study[13] described BC1-BC5,
221 each with specific insertion sites of IS elements. Any novel insertion sites identified in

222 this present study were therefore numbered from 'BC6' forward. The novel IS*Efa16,*
223 IS*Efa17,* IS*Efa18* sequences were identified in this study and registered in the
224 ISFinder public database[27].
225
226 **Accession numbers.** DNA sequences from this study have been deposited under
227 Bioproject PRJNA682584. Individual accession numbers for all sequence files are
228 provided in **Dataset S1A**.
229
230 **Results**
231 **VREfm isolates belonging to ST17 and ST18 were prevalent in the collection.**
232 Genome sequencing of 46 VREfm isolates and one vancomycin-sensitive comparator
233 collected from rectal surveillance of high-risk patients at a Dallas, Texas, hospital was
234 performed with Oxford Nanopore MinION and Illumina technologies. High quality
235 genome sequences were generated by hybrid assembly of the sequencing reads from
236 these two platforms. 31 closed and 16 draft genome assemblies were generated
237 (**Table S1**). In our previous analysis of these isolates using PCR, all isolates were
238 found to be *vanA*-positive and *vanB*-negative by PCR, with the exception of the
239 isolate 163-1, which had neither *vanA* or *vanB* and had a vancomycin MIC of 2 μg/mL
240 by broth microdilution[16]. We included 163-1 in this study for comparative purposes to
241 evaluate its relationship to the *vanA*-positive isolates in the collection. All 46 *vanA*-
242 positive isolates sequenced in this study grew with 256 μg/mL vancomycin in the
243 growth medium.
244
245 Phylogenetic analysis based on alignment of 1706 core genes allowed us to assess
246 the epidemiology and diversity of the VREfm isolates (**Figure 1**). Two reference

247 genomes, *E. faecium* 1,231,501[6] and *E. faecium* ATCC® 700221™[28], were included
248 as representatives of *E. faecium* clade A2 and A1, respectively. *E. faecium* 1,231,501
249 is a human bloodstream isolate from the USA and is susceptible to vancomycin. *E.*
250 *faecium* ATCC® 700221™ is a USA isolate and VanA-type VREfm control strain used
251 in antimicrobial susceptibility testing[29]. As expected, the Dallas VREfm isolates
252 clustered with the clade A1 reference *E. faecium* ATCC® 700221™, although a high
253 degree of branching within this clade was observed, revealing genetic diversity within
254 the collection.
255
256 MLST is commonly used to study relationships among VREfm isolates worldwide, and
257 this analysis provides the opportunity to compare the Dallas isolates with other
258 VREfm strain collections. A total of 10 different sequence types (ST), all belonging to
259 clonal complex 17 (CC17), were observed among the collection (**Figure 1, Figure 2**).
260 Isolates belonging to ST17 and ST18 are the most prevalent in the collection,
261 comprising 50% of the isolates, which clustered together in the core genome tree
262 (**Figure 1, Figure 2**). Six genotypes (ST18, ST664, ST1703, ST736, ST80 and
263 ST612) are double locus variants of ST17. The ST262 and ST412 are triple locus
264 variants and ST1383 is a single locus variant of ST17 (**Dataset S1C**). ST1703, a
265 double locus variant of ST17 (*pstS* allele-20*;* novel *gyd* allele-70), is a novel ST in the
266 MLST database and comprised 11% of the collection. Isolate 163-1 branched away
267 from the rest of the isolates and belonged to ST262, which is still within CC17. Taken
268 together, these data indicate that *E. faecium* CC17 with varied genotypes colonized
269 Dallas patients at the time of the study.
270
271 **Large numbers of plasmids are present in the Dallas isolates**. The plasmid
272 content of the Dallas isolates was analyzed by an *in silico* detection method utilizing

273 the conserved nucleotide sequences of replication initiation (*rep*) genes[23]. Of the 259
274 closed circular putative plasmids, 141 were assigned to 10 known *rep* families
275 described in the existing PlasmidFinder database. The remaining 118 putative
276 plasmids were categorized into 9 novel plasmid groups (pMI1 to pMI9) and 8 other
277 elements (discussed in the next section).
278
279 The Dallas isolates harbored a variable number of plasmids (n=3 to 12) of sizes
280 ranging from 1.9 to 294 kb (**Figure 1; Dataset S1D**). The most common elements
281 were a megaplasmid (126 to 294 kb) and a pRUM-like plasmid (17 to 63 kb), which
282 were each present in all isolates except 144-1, which lacked the megaplasmid. Both
283 the reference strains (*E. faecium* 1,231,501 and ATCC® 700221™) harbored the
284 megaplasmid, but 1,231,501 lacked the pRUM-like plasmid. Each of the
285 megaplasmids harbored the *rep* gene *repUS15*, encoding a RepA_N protein motif.
286 The pRUM-like plasmid harbored *rep17* or both *rep17* and *rep18*, which we refer to as
287 a mosaic plasmid. Isolates collected from the same patient had identical plasmid
288 profiles in 2 cases (patients 124 and 9) while different plasmid profiles were observed
289 in 3 other cases (patients 100, 71, and 16).
290
291 To determine whether plasmid replicons were enriched among specific lineages, the
292 percentage presence/absence of specific *rep* genes was quantified among the
293 isolates belonging to 4 STs (ST17, ST18, ST1703 and ST664) that included ≥5
294 isolates. The other 6 STs (ST262, ST1383, ST612, ST412, ST80, and ST736) were
295 not included in this analysis because they were represented by <5 isolates (**Dataset**
296 **S1E**). Chromosomally integrated *rep* genes *repUS12* and *repUS43,* encoding the
297 protein motifs Rep_1 and Rep_trans, were observed in 45% and 22% of the isolates,
298 respectively, and were absent from isolates belonging to ST17, ST18 and ST1383

299 (**Dataset S1E**). The *rep14* family was present in 80% and 90% of ST17 and ST18
300 isolates, respectively, but was absent in ST1703 and ST664 isolates. The *rep18*
301 family plasmid was present in ST17 (80%) and ST664 (100%) isolates but absent in
302 ST1703 (0%) and rare in ST18 (30%). In summary, the *rep* genes *repUS12, repUS43*,
303 *rep18* and *rep14* had an ST-specific distribution among the isolates. No conclusion
304 could be drawn regarding other *rep* genes and STs due to low sample size.
305
306 **Uncategorized plasmids were sorted into 9 plasmid groups, pMI1 to pMI9.**
307 PlasmidFinder[23] did not identify *rep* genes in a total of 118 circular entities. Of these
308 118 circular entities, 110 were established to be plasmids based on the presence of
309 confirmed *rep* genes not currently represented in the PlasmidFinder database (see
310 Materials and Methods). The remaining 8 circular entities were designated as putative
311 excision products of genomic islands (**Supplementary text, Dataset S1F-G**)
312 including a 5 kb *vanA*-encoding circular fragment of Tn*1546* in isolate 111 (discussed
313 further in a later section).
314
315 The 110 plasmids not classifiable by PlasmidFinder were sorted into 9 plasmid
316 groups, pMI1 to pMI9, based on phylogenetic analyses of the *rep* genes (**Figure S1;**
317 **Dataset S1H**). The nucleotide sequences of the *rep* genes and their corresponding
318 amino acid sequences shared >95% identity within each pMIX plasmid group, except
319 for pMI4 (**Table 1**). The clustering of the *rep* gene sequences belonging to the pMI4
320 plasmid group supported their inclusion in a single group, despite lower pairwise
321 sequence identities (**Figure S1**). A total of 6 Rep protein families (RepA_N, Rep_3,
322 Rep_1, Rep_trans, Rol_rep_N, and Rep_2) were encoded by the *rep* genes of the 9
323 pMIX groups (**Dataset S1H**). Specifically, the *rep* genes of pMI1 and pMI3 encoded
324 protein families RepA_N and Rep_1, respectively. The *rep* genes of pMI2, pMI5, pMI6

325 and pMI7 encoded the protein family Rep_3. The *rep* genes of pMI4 and pMI8
326 encoded protein families Rep_trans and Rep_2, respectively. The *rep* of pMI9 was
327 the only exception, encoding two Rep protein families, Rep_trans and Rol_rep_N
328 (**Table 1**).
329
330 Each of the Dallas isolates harbored 1 to 5 of the pMIX plasmid groups, except the
331 isolate 83 (ST17), which had none of them (**Figure 1**). Neither of the two reference
332 strains harbored any of the pMIX plasmid groups. The pMI1 was the most frequently
333 occurring pMIX plasmid group, identified in 75% of the isolates (**Table 1**). The pMI8
334 and pMI9 were the rarest ones, found in only isolates 51_1 (ST612) and 5 (ST18),
335 respectively. The pMIX plasmid groups were not enriched in any specific ST.
336 Plasmids within a very narrow size range (standard deviation 0.2 to 2.3) were
337 grouped together (with the exception of pMI1), although plasmid size was not taken
338 into consideration in assigning plasmid groups. The pMI1 consisted of plasmids size
339 ranging from 23 to 62 kb (standard deviation 11.2).
340
341 To summarize, 9 new plasmid groups, pMI1 to pMI9, were defined for 110 plasmids
342 based on *rep* gene sequences that encoded 6 different Rep protein families. Most of
343 the Dallas isolates harbored 1 to 5 plasmids belonging to these newly defined plasmid
344 groups.
345
346 **Resistance genes were encoded in both chromosomes and plasmids**. Different
347 combinations of 12 resistance genes providing resistance against 5 classes of
348 antibiotics (aminoglycosides, glycopeptides, macrolides, tetracycline, and
349 trimethoprim) were identified among the Dallas isolates (**Figure 3**; **Dataset S1I;**
350 **Supplemental text**). The most commonly occurring resistance genes were *aac(6')I-1*

351 (100% of isolates), *msr* (100%), *vanA* (95%), *tetM* (95%), *ermB* (85%), and *aph(3')*
352 (85%). These resistance genes were encoded in the chromosome and mostly in two
353 plasmid types, the pRUM-like and the megaplasmid. The pRUM-like plasmids
354 predominantly encoded 4 of the resistance genes (*ant(6)-Ia, aph(3')III, vanA,* and
355 *ermB*) conferring resistance against aminoglycosides, vancomycin and erythromycin
356 (**Figure 3**). The *repUS7* plasmid (pMG1) was found among only 3 of the Dallas
357 isolates (66 (ST18), 144-1 (ST18) and 51-4 (ST612)) and encoded *dfrG,* a less
358 common resistance gene among the Dallas isolates, conferring resistance against
359 trimethoprim.
360
361 All of the Dallas isolates, except for 163-1, encoded VanA-type vancomycin
362 resistance genes, where *vanA* shared 99.9% (nucleotide) and 99.7% (amino acid)
363 sequence identity with the reference sequence (GenBank AAA65956.1). The
364 reference sequence is encoded within the transposon Tn*1546* (GenBank M97297),
365 originally described in the human fecal isolate *E. faecium* BM4147 collected from
366 France in 1988[30] (**Dataset S1I)**. A classical structure of the *vanA* gene cluster
367 consisting of *vanHAX* and *vanYZ* encoded within Tn*1546* elements was observed
368 among the Dallas VREfm isolates, except for the isolate 17-1, which lacked the genes
369 *vanY* and *vanZ*. The Tn*1546* structure varied among the isolates in terms of insertion
370 site (chromosome versus plasmid, IS elements, and point mutations, discussed
371 further below.
372
373 **The Tn*1546* is borne on pRUM-like plasmids in the majority of isolates**. The
374 complete Tn*1546* was carried by the pRUM-like plasmid in 42 of 46 *vanA*-positive
375 isolates (**Figure 4**). The isolate 111 did not encode the entire Tn*1546* in the pRUM-
376 like plasmid; rather, it encoded *vanRS* in a 17 kb pRUM-like plasmid while *vanHAX*

377 was encoded within a separate 5 kb circular element (te111_5kb) (**Dataset S1F**).
378 Tn*1546* was chromosomally encoded in 3 isolates, 53-1, 53-2 and 91, each belonging
379 to ST17, despite all 3 harboring a pRUM-like plasmid (**Table 2**).
380
381 Two toxin-antitoxin (TA) systems, *axe-txe* ($TA_{axe\text{-}txe}$) and a *relE/parE* family toxin
382 associated with an uncharacterized antitoxin component (TA_{relE}), were observed
383 alone or simultaneously on the pRUM-like plasmids (**Table 2**). Both of the TA systems
384 were actively expressed in two representative VREfm isolates (**Figure S2**).
385
386 The 43 pRUM-like plasmids encoding Tn*1546* or a fragment of Tn*1546* (for isolate
387 111) were categorized into the following four groups based on the presence or
388 absence of replication (as categorized by PlasmidFinder) and stability modules on
389 their backbones (**Table 2, Figure 5**). A previous study utilized the term "pRUM
390 derivatives" where plasmids were characterized by analyzing *rep* and TA systems[12].
391
392 i) Group1 ($rep_{17;}$ rep_{18}; $TA_{axe\text{-}txe;}$ TA_{relE}): The pRUM-like plasmids categorized
393 into group 1 harbored a mosaic replication module $rep_{17/18}$ as well as both
394 of the stability modules $TA_{axe\text{-}txe}$ and TA_{relE}. All the group 1 plasmids
395 encoded the *rep18b* (accession number CP018068), except isolate 113,
396 which encoded the *rep18a* (accession number AB158402). A total of 16 of
397 the 43 plasmids belonged to this group. The plasmids had a narrow size
398 range (44 to 48 kbp) and were carried by VREfm isolates from 4 genetic
399 backgrounds (ST17, ST18, ST1383 and ST80).
400 ii) Group 2 (rep_{17}; $TA_{axe\text{-}txe}$; TA_{relE}): A total of 17 of 43 plasmids harbored the
401 replication module rep_{17} and both of the stability modules $TA_{axe\text{-}txe}$ and
402 TA_{relE}, and were categorized into group 2. The group 2 plasmids could be

403 classified further based on size and genetic background of the isolates
404 carrying them. Group 2a plasmids (29-45 kb) were carried by isolates
405 belonging to ST18 (5, 66, 97-1, 124-1, 124-2, 124-3, and 162), ST80 (16-
406 1, 16-2), ST17 (93-2), and ST612 (51-4), whereas larger (57-63 kb) group
407 2b plasmids were present in all 6 isolates belonging to ST664 (51-1, 87-1,
408 93-1, 93-2, 155, and 160). The reference strain ATCC® 700221™
409 harbored the group 2a pRUM-like plasmid. Sequence analysis revealed
410 that Group 2b plasmids possess a set of genes that are specific to this
411 group and present in no other group (**Dataset S1J**), including a ~7.5 kb
412 region encoding helix-turn-helix XRE-family like proteins (cd00093)
413 inserted immediately upstream of the *van* gene cluster (**Figure S3**).

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

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

424

425 It is interesting to note that the isolate 163-1, which belongs to CC17 but lacks
426 vancomycin resistance genes, harbored a 25 kb pRUM-like plasmid. This pRUM-like
427 plasmid did not harbor any part of the Tn*1546* and did not belong to any of the
428 plasmid sub-groups defined above. The presence of the modules *rep17* and $TA_{axe\text{-}txe}$

429 and the absence of both of the modules *rep18* and TA_{relE} was unique among the
430 Dallas isolates but was described before in the originally described pRUM plasmid in
431 2003[31].
432
433 The robustness of the *rep*/TA typing system was confirmed using two approaches: a
434 core genome phylogeny based on sequence variation in 7 pRUM-like core genes
435 (*rep17,* IS*1216, vanS,* ImpB/MucB/SamB family protein, hypothetical protein, plasmid
436 partitioning protein, and replication control protein PrgN) (**Figure S5**), and an analysis
437 of gene presence/absence across 40 pRUM-like plasmids using Roary[19] **(Dataset**
438 **S1J).** The Group 2a pRUM-like plasmid from the reference strain *E. faecium*
439 ATCC® 700221™ was included in the analysis; while 4 of the pRUM-like plasmid
440 sequences (71-2, 100-2, 124-2, and 124-3) were not included due to sequence
441 redundancy. The pRUM-like plasmids were clearly divided into two branches in the
442 phylogenetic tree (**Fig S8**). Interestingly, these two groups corresponded to the
443 presence/absence of the $TA_{axe\text{-}txe}$. Members of the plasmid groups 1, 2a and 2b
444 clustered together, while those of group 3 and 4 were placed into a separate branch.
445 The gene presence/absence heatmap (**Dataset S1J**) of 40 pRUM-like plasmids
446 clearly identified genes specific to certain plasmid groups, other than the two stability
447 and replication modules.
448
449 **Dallas VREfm harbored 17 variants of Tn*1546*.** The Tn*1546* region of the Dallas
450 isolates varied from the typical structure consisting of ORF1 (transposase), ORF2
451 (resolvase) and *van* genes flanked by IR_L (inverted repeat, left) and IR_R (inverted
452 repeat, right) (**Figure 6**). A total of 17 different sequence variations in the Tn*1546*
453 were observed among 45 of the Dallas isolates (**Figure 6, Tables 3-4**). These 17
454 variations were classified using a nomenclature system developed previously[13] based

455 on the presence of IS elements, point mutations, and deletions relative to the
456 reference Tn*1546*. VREfm belonging to the same Tn*1546* group shared 99.8%-100%
457 nucleotide sequence identity in the Tn*1546* region. The isolate 111 was excluded
458 from this grouping since it harbored a fragmented Tn*1546* in a pRUM-like plasmid and
459 a circular entity (**Dataset S1F, Figure S2**). The Tn*1546* variation was mostly
460 associated with the distribution of 7 different IS elements (IS*1251*, IS*1216*, ISEfa*16*,
461 IS*Efa5*, IS*Efa17*, IS*256*, IS*Efa18*), three of which are novel to the Dallas isolates.
462 Most of the Tn*1546* groups described in this study are novel, except for group BC2,
463 which was reported previously in Poland[13]. The most predominant Tn*1546* groups
464 were BC9 (n=14) and J (n=6). The BC9 variety was observed in ST17 (n=9) and
465 ST18 (n=5) isolates. The reference strain ATCC® 700221™ harbored the C2
466 structural variant of Tn*1546* that was described previously[13] but is not present in our
467 Dallas VREfm isolates.
468
469 Identical Tn*1546* groups were observed for isolates collected from the same patient
470 (for example, all three VREfm isolates from patient 124 harbored Tn*1546* type BCCL),
471 except patients 51 and 93 (**Table 3**). Isolates 51-1 and 51-4 harbored Tn*1546* type J
472 and BC2 on the plasmid backbones 2b and 2a respectively. Isolates 93-1 and 93-2
473 harbored the same Tn*1546* variety J on pRUM-like plasmid group 2b while 93-3
474 harbored BCM on plasmid group 2a. Interestingly the 51-1 (ST664) and 51-4 (ST612)
475 isolates belonged to different clades in the phylogenetic tree. The isolates 93-1
476 (ST664) and 93-2 (ST664) clustered together but 93-3 (ST17) was placed in a
477 separate clade. It is possible that pRUM-like vancomycin resistance plasmids were
478 acquired by different *E. faecium* genotypes in two separate events in each of these
479 patients.
480

481 Specific structural variations of the Dallas Tn*1546* elements are described below.

482

483 **The 3' but not 5' boundary of the Tn*1546* was conserved among the Dallas**
484 **isolates.** Each of the 17 Tn*1546* groups preserved the IR_R boundary of Tn*1546* but
485 lacked the IR_L boundary (**Figure 6**). The only exception was the isolate 17-1 (group
486 BBCD) that lost both of the IR boundaries due to insertion of IS*1216*. The complete or
487 partial deletion of ORF1 (transposase) and ORF2 (resolvase) was observed in all of
488 the Tn*1546* groups. This might decrease the likelihood of these Tn*1546* to
489 disseminate to other isolates without a plasmid as a vehicle.

490

491 **IS*1251* and IS*1216* were the most common IS elements.** IS*1251* was present in
492 84.4% (38 of 45) of the isolates (**Table 3**). It was inserted at Tn*1546* position 5820 in
493 the *vanS*-*vanH* intergenic region, in reverse orientation to the *van* genes (**Figure 6**).
494 The IS*1251* generated an 8 bp direct repeat (5'-AAAATTAT-3') in the regions where it
495 was inserted. IS*1216* was inserted at either the 5' end of Tn*1546* in 71.1% (32 of 45)
496 or in the intergenic region of *vanX*-*vanY* in 17.8% (8 of 45) of the isolates.

497

498 **The *van* genes and *vanR* and *vanH* promoters were intact in most isolates,**
499 **despite IS element insertions.** IS*1251* insertion at Tn*1546* position 5820 did not
500 disrupt the two 12 bp VanR binding sites located at positions 5846 and 5879 (**Figure**
501 **S4**). The 70 bp VanR-interacting region (-110 to -41) from positions 5935 to 6004 is
502 also intact. The 197 bp long proposed *vanR* promoter region (-173 to +24) is located
503 at the position 3,803 to 3,999 bp of the Tn*1546*. The isolate 137 (Tn*1546* group BCK)
504 has an IS*Efa16* inserted at the position 3664 to 3763 that was responsible for a 99 bp
505 deletion of resolvase at 3' region but the *vanR* promoter region remained
506 uninterrupted. The *vanR* promoter region including the 12 bp consensus VanR

507 binding site is disrupted in the isolates 51-1, 87-1, 93-1, 93-2, 155, 160 from group J
508 because of the insertion of IS*1216* at position 3914. The *vanR* promoter region for
509 other isolates remained uninterrupted.
510
511 Deletion of *vanR* and *vanS* was observed in the isolate 91 (variant group D1).
512 Expression of the *vanHAX* genes may be constitutive instead of inducible in this
513 isolate. No mutations were identified in *vanR* and *vanS* in the rest of the groups.
514
515 **Two point mutations in *vanHAX* were fixed in the population.** All isolates
516 harbored a point mutation (T7658C) in *vanA* that generated a valine to alanine
517 substitution (**Table 3**) and a second point mutation (G8234T) in *vanX* that generated a
518 lysine to asparagine substitution. 62.2% (28 of 45) of the isolates harbored a third
519 point mutation (C9692T) in *vanY* that was also present in the reference strain
520 ATCC® 700221™. This point mutation generates a proline to leucine substitution in
521 VanY. Isolates 53-1 and 53-2 (group BC10) had point mutations C6513A and G6514A
522 in *vanH* that generated an arginine to lysine substitution. Isolate 144-1 harbored a
523 unique G9566A point mutation in *vanY* that conferred a glycine to glutamic acid
524 substitution. No mutations in *vanZ* were observed in any isolates.
525
526 **Discussion**
527 This study provided a snapshot of fecal VREfm genotypes among hospitalized
528 patients in Dallas, Texas. To our knowledge, this was the first whole-genome
529 epidemiological analysis of fecal VREfm isolates from patients without an active
530 infection due to VREfm in this area.
531

532 Genetic diversity was observed among all of the 47 isolates analyzed. The most
533 prevalent VREfm isolates, comprising 50% of the population, belonged to ST17 and
534 ST18. These STs were previously reported to be responsible for outbreaks in
535 countries including Portugal[12], Spain[32], Ireland[33], Denmark[34], Columbia[35], Brazil[36],
536 Canada[37]. The other 8 STs, identified among our collection (ST1383, ST612, ST412,
537 ST80, ST1703 and ST664), were single, double, or triple locus variants of ST17. The
538 ST1703 was a novel sequence type described in this study. Notably, ST182, which
539 caused an outbreak in San Antonio, Texas in the 1990s[38], was not detected in our
540 collection. The isolate 163-1 was identified as a presumptive VREfm from clinical
541 collection and was co-isolated with a *vanA*-positive VREfm not sequenced in this
542 study[16]; but was sensitive to vancomycin by broth microdilution and negative for *vanA*
543 and *vanB* by PCR screening[16]. This isolate was incorporated in our study despite
544 being sensitive to vancomycin because it provided a reference point to compare with
545 other Dallas isolates. This isolate generated its own clade, separate from the *vanA*-
546 positive isolates in the phylogenetic tree, and its genetic background, ST262, has not
547 been reported as VREfm to our knowledge. However, interestingly, the ST262 is a
548 triple locus variant of ST17 (*pstS-7, atpA-7,* and *gyd-5*) and still belongs to the CC17.
549 To conclude, VREfm isolates belonging to a combination of previously described and
550 new genetic backgrounds colonize Dallas, Texas, patients.
551
552 The number of core genes identified among the Dallas isolates was 1706, which is
553 much higher than the previously reported 1344[5] and 955[39] core genes described by
554 other groups. The lower number of samples (47 isolates) collected from a specific
555 source (hospitalized, high-risk patients with fecal VREfm colonization from Dallas,
556 Texas) is likely responsible for the larger number of core genes in our current study.
557 The previous studies utilized larger sample sizes (73[5] and 1,644[39]), were collected

558 from different sources including hospitalized patients, healthy human volunteers, and
559 animals, and included both VREfm and non-VREfm.
560
561 Completely closed plasmid sequences were achieved in this study by producing high
562 quality genome assemblies. This made it possible to perform comparative analyses of
563 entire plasmid sequences, which is a significant advantage compared to many
564 previously employed techniques to analyze resistance plasmid diversity, including
565 PCR-based typing schemes, restriction fragment length polymorphism (RFLP)
566 comparisons, and hybridization[12].
567
568 All the plasmid-borne Tn*1546* in Dallas isolates were encoded in pRUM-like plasmids
569 that were clustered into 4 subgroups (1, 2a, 2b, 3, and 4), defined by the presence of
570 the replication module *rep17* and a combination of other replication and stability
571 modules. Speculatively, deletion or insertion of $TA_{axe\text{-}txe}$ drove the evolution of the
572 group 3 and 4 pRUM-like plasmids. The vancomycin-sensitive isolate 163-1 also
573 harbored a pRUM-like plasmid, without Tn*1546,* that did not belong to any of the sub-
574 groups described in this study. This striking homogeneity in the plasmid backbone
575 encoding *vanA* was surprising, because previous studies identified a mixture of
576 different plasmids (pRUM-like, Inc18, mega-plasmid pLG1) as being responsible for
577 vancomycin resistance in clinical isolates collected from a certain geographic region[12].
578 However, our study differs from these previous studies in temporal and geographic
579 features of sample collection. In previous studies, samples were collected over 2 to
580 12 years[12,13,40] and from different regions of countries including Portugal[12],
581 Poland[13]and Germany[40].
582

583 The patients from whom the isolates in this study were collected resided in different
584 areas of Dallas and in unconnected facilities including nursing homes and personal
585 residences. The similarities among them were that they were high-risk patients,
586 having recently been in a healthcare environment or having antibiotic exposure.
587 Therefore, the presence of only one type of plasmid backbone carrying *vanA* in the
588 Dallas isolates is striking. However, the existence of a common source of VRE acting
589 as a reservoir and transferring a vancomycin resistance plasmid to *E. faecium* of
590 different genetic background is an unlikely scenario since the pRUM backbone
591 sequences and Tn*1546* structure varied.
592
593 The phylogenetic analysis of pRUM-like plasmids in this study suggested an
594 evolutionary event driven by either insertion or deletion of the $TA_{axe\text{-}txe}$ which was
595 originally described in the pRUM plasmid[31]. pRUM-like plasmids lacking $TA_{axe\text{-}txe}$ were
596 reported previously[12] although the possible implication of absence of $TA_{axe\text{-}txe}$ on
597 pRUM-like plasmid sub-grouping was not taken into consideration. The current study
598 suggests a possibility of evolution of the pRUM-like plasmids by deletion of the TA
599 system *axe-txe* since they have a second active TA_{relE}. pRUM-like plasmid groups 3
600 and 4, defined in this study, may be descendants of groups 1 and 2 in Dallas-area
601 VREfm (**Figure S5**). Whether this impacts the maintenance of these plasmids in the
602 absence of antibiotic selection, or their transfer efficiencies, requires further
603 experimental study.
604
605 A total of 17 variations of the Tn*1546* region, defined by the presence of different IS
606 elements, point mutations, and deletions, were observed among the Dallas isolates.
607 In absence of a strict guideline for the nomenclature of the Tn*1546* variants, a recent
608 one described in 2017[13] was followed in this study. A slight variation in Tn*1546*

609 nomenclature was utilized in 2016 where variants with IS*1251* were designated as
610 "type F" or "US hospital type"[9], which are designated as group "C" in the current
611 study. Three newly identified IS elements (IS*Efa16,* IS*Efa17,* and IS*Efa18*) within the
612 Tn*1546* were observed in this study. Moreover, two point mutations within Tn*1546*,
613 T7658C and G8234T, which were ubiquitous among the Dallas VREfm isolates, were
614 reported previously but never in the same Tn*1546* group[13]. The point mutation
615 T7658C was reported in the Tn*1546* groups BC5, BC1, BC2, BC3, and BC4 while
616 G8234T was reported in groups A2, C1, C2, BB1, BB2, BC1, BC2, BC3, BC4, BC5,
617 BBBI1, and BBBI2[13].
618
619 The Tn*1546*-borne vancomycin resistance genes were carried by pRUM-like
620 plasmids, which was categorized into plasmid groups 1, 2a, 2b, 3, and 4 based on
621 replication and stability modules. Both of the pRUM-like plasmid groups 1 and 2a
622 harbored the Tn*1546* variation BC9 (**Figure 7**). This was the only example of two
623 different plasmid groups carrying the same Tn*1546* variation. Therefore, the plasmid
624 backbone may have diverged into groups 1 and 2a after acquisition of the BC9 type
625 Tn*1546*. The same variation of Tn*1546*-type transposon was never observed for the
626 plasmid groups 2b, 3 and 4. Therefore, an independent evolution of Tn*1546* and the
627 plasmid backbone can be suggested for the rest of the resistance plasmid groups.
628
629 The resistance plasmid group 1 harboring the same Tn*1546* variant, BC9, in two
630 different genetic backgrounds, ST17 (isolate 1, 20, 57, 83, 103, 107-1, 131-2, 148,
631 and 161) and ST18 (isolate 100-1 and 100-2), was observed. Horizontal transfer of
632 this specific plasmid between ST17 and ST18 is possible (**Figure 7**). Similarly,
633 horizontal transfer of plasmid group 1 with Tn*1546* variant BC11 between isolates
634 belonging to ST18 and ST1383 may have occurred, as well as horizontal transfer of

635 plasmid group 2a with Tn*1546* variant BC2 between isolates belonging to ST80 and
636 ST612 (**Figure 7**). No other evidence of horizontal transfer of pRUM-like plasmids
637 among different VREfm genetic backgrounds was observed. Different plasmid groups
638 present in isolates with the same genetic background (group 1, 2a in ST17, ST18,
639 ST80; group 3, 4 in ST1703) suggested independent acquisition events of the
640 resistance plasmids within those STs. The resistance plasmid groups 1 and 2a were
641 observed in the common STs 17 and 18. The other resistance plasmid groups 2b, 3
642 and 4 were carried by the less prevalent STs. So, a combination of both horizontal
643 transfer and independent acquisition events followed by IS element-driven Tn*1546*
644 evolution shaped the VanA-type VREfm population in this area.
645
646 To conclude, the molecular diversity of multidrug resistant VREfm isolates in Dallas,
647 Texas, collected in the year 2015 was documented. VREfm with a novel genetic
648 background, ST1703, was reported in this study. No clonal spread of any particular
649 VREfm genotype could be detected in this isolate collection. However, strikingly,
650 vancomycin resistance genes were borne by pRUM-like plasmids in almost all
651 isolates despite the diversity of patient characteristics. Diversity of Tn*1546* was
652 described and new variations involving 3 novel IS elements (IS*Efa16,* IS*Efa17* and
653 IS*Efa18*) were also described among the VREfm collection. Horizontal transfer of
654 resistance plasmids between ST17-ST18 and ST80-ST612 was speculated since
655 VREfm of these genetic backgrounds carried the same variation of Tn*1546* on the
656 same plasmid backbone. Independent acquisition of resistance was observed in the
657 rest of the isolates. Therefore, it was concluded that VREfm representatives from
658 different sources carrying unique Tn*1546* generated the demography of VREfm in
659 Dallas, Texas.
660

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. Multidrugresistant 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 apecified 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 bloodstreasolates 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*

791 *faecium* during its emergence in a city in southern Texas. *Clin. Infect. Dis. an*
792 *Off. Publ. Infect. Dis. Soc. Am.* **21**, 1234–1237 (1995).
793 39. Arredondo-Alonso S, Top J, McNally A, Puranen S, Pesonen M, JPensar
794 J, Marttinen P, Braat JC, Rogers MRC, van Schaik W, Kaski S, Willems
795 RJL, Corander J, Schürch AC. Plasmids Shaped the Recent Emergence of the
796 Major Nosocomial Pathogen *Enterococcus faecium*. *MBio* **11**, (2020).
797 40. Bender JK, Kalmbach A, Fleige C, Klare C, Fuchs S, Werner G. Population
798 structure and acquisition of the vanB resistant determinant in German clinical
799 isolates of *Enterococcus faecium* ST192. *Sci. Rep.* **6**, (2016).
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816

817 **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	*rep* gene information				
					rep gene size (bp)	% identity within group (*rep* DNA)[3]	% identity within group (*rep* aa)[4]	*Rep* 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	Rol_Rep_N	PF18106.1
								Rep_trans	PF02486.19

818

819 [1]Specifies the total number of isolates among the 47 Dallas isolates harboring the
820 particular pMIX group plasmids.

821 [2]The % of isolates harboring the specific plasmid group.

822 [3,4]The "% identity (DNA)" and the "% identity (aa)" specifies the % identical sites
823 calculated from the MUSCLE sequence alignment of DNA and amino acid of the *rep*
824 genes belonging to the same pMIX group, respectively.

825 [5,6] The *rep* gene encoded protein family (pfam) and the corresponding pfam
826 accession numbers, respectively.

827

828

829

830

831

832

833

834

835

836 **Table 2. Features of the pRUM-like plasmid groups**

plasmid groups	MLST[1]	Isolate	*rep*[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 [1]Multiple locus sequence type (MLST) of the isolates

839 [2,3]The presence of *rep* and toxin-antitoxin (TA) genes are indicated by blue boxes.
840 The pRUM-like plasmid from isolate 113 harbored *rep18a* (reference gene Accession
841 no. AB158402) and is symbolized as “a” in the corresponding box. The *rep18* for rest
842 of the resistance plasmids were *rep18b* (reference gene Accession no. CP018068).
843 [4]The average size of the plasmids for each plasmid group type is shown along with
844 the standard deviation (Avg. ± stdev).
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864

865 **Table 3. Features of the Tn*1546* groups observed among the Dallas isolates**

pRUM-like plasmid group	Tn*1546* Group	Isolate	Identity within Tn*1546* group[1]	No. of Isolates in Tn*1546* 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	IS*1216*, IS*1251*, IS*Efa17*					
	BC8	113	N/A	1	IS*1216*, IS*1251*					
	BC11	71-1	99.80%	3	IS1216, IS1251					
		71-2								
		144-1								
	BC9	1	99.80%	14	IS*1216*, IS*1251*					
		20								
		57								
		83								
		100-1								
		100-2								
		103								
		107-1								
		131-2								
		148								
		161								
2a		5								
		66								
		162								
	BC2	16-1	100%	3	IS*1216*, IS*1251*					
		16-2								
		51-4								
	BBC	97-1	N/A	1	IS*1216*, IS*1251*					
	BCM	93-3	N/A	1	IS1216, IS1251, IS*Efa18*					
	BCCL	124-1	100%	3	IS*1216*, IS*Efa17*, IS*1251*, IS*1251*					
		124-2								
		124-3								
2b	J	51-1	100%	6	IS*256*					
		87-1								
		93-1								
		93-2								
		155								
		160								
3	C3	52	100%	2	IS*1251*					
		55								
	BC6	9_1	100%	2	IS*1251*, IS*1216*					
		9_2								
	BC7	121	N/A	1	IS*1251*, IS*1216*					
4	BCB	154-1	100%	2	IS*1216*, IS*1251*, IS*1216*					
		158								
	BCK	137	N/A	1	IS*Efa16*, IS*1251*, IS*1216*					
	BBCD	17-1	N/A	1	IS*1216*, IS*Efa5*, IS*1251*, IS*1216*					
Chromosome integrated Tn*1546*	D1	91-1	N/A	1	IS*Efa5*					
	BC10	53-1	100%	2	IS1216, IS1251					
		53-2								

866

867 [1] Nucleotide sequences of the Tn*1546* region for the Tn*1546* groups, including more

868 than one isolate, were aligned and % identity was determined.

869 [2]The name of the IS elements present in the Tn*1546* group are listed.

870 [3]Point mutations in the van gene cluster relative to the reference Tn*1546* (accession

871 number M97297). Presence of a point mutation is represented by a colored box.

872

873

874

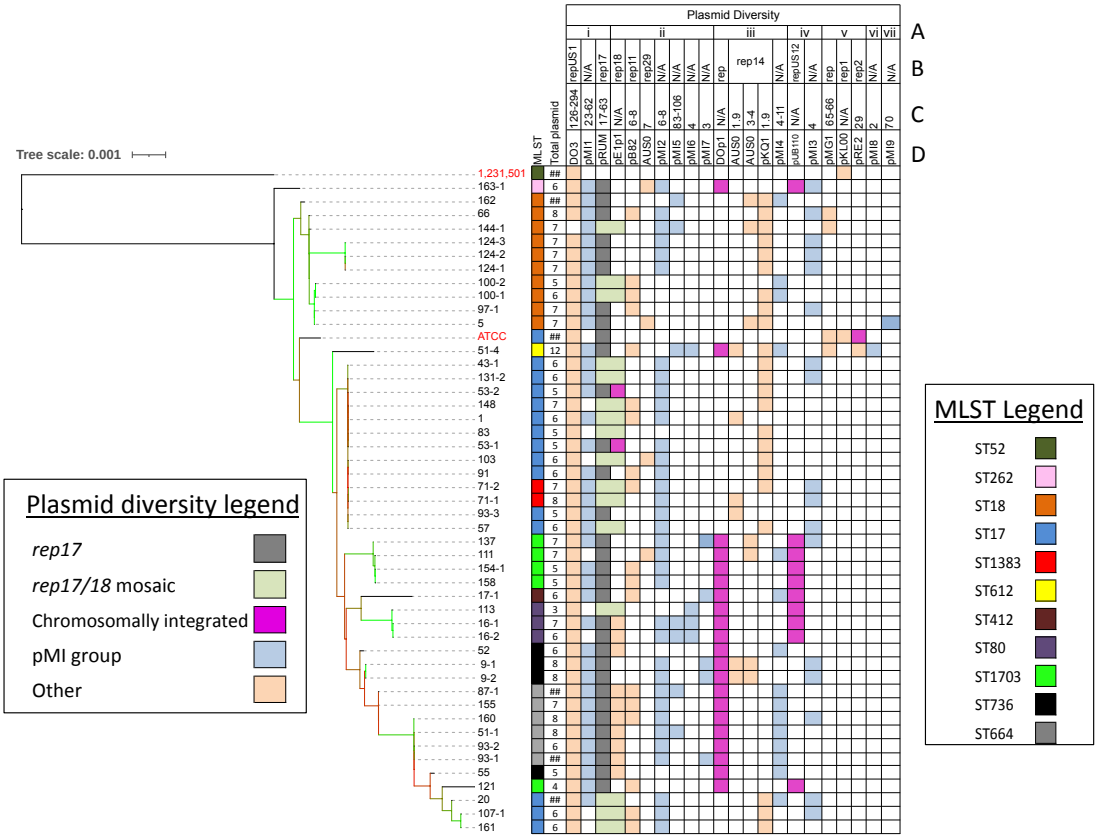


Figure 1. Plasmid content of Dallas *E. faecium* isolates. The phylogenetic tree of 49 *E. faecium* strains is based on an alignment of SNPs in 1706 core genes. The branches colored green, golden and red are supported by a bootstrap value of >70, 40-60, and <30 respectively. The reference genomes *E.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

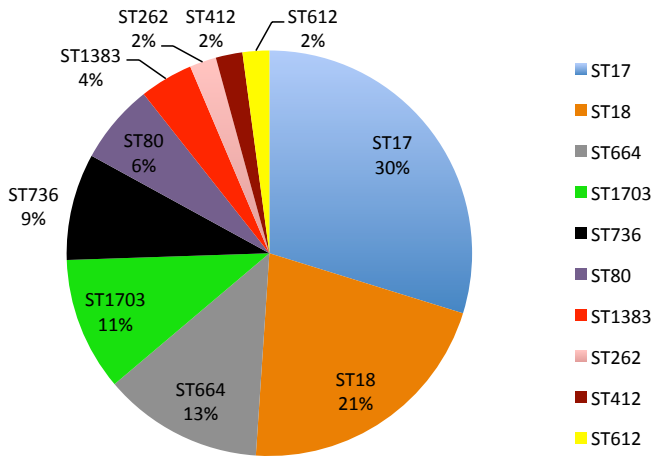


Figure 2: Sequence type (ST) distribution of Dallas *E. faecium* isolates. A total of 10 STs were observed across 47 isolates. The frequency of STs (percent of total isolates) is shown.

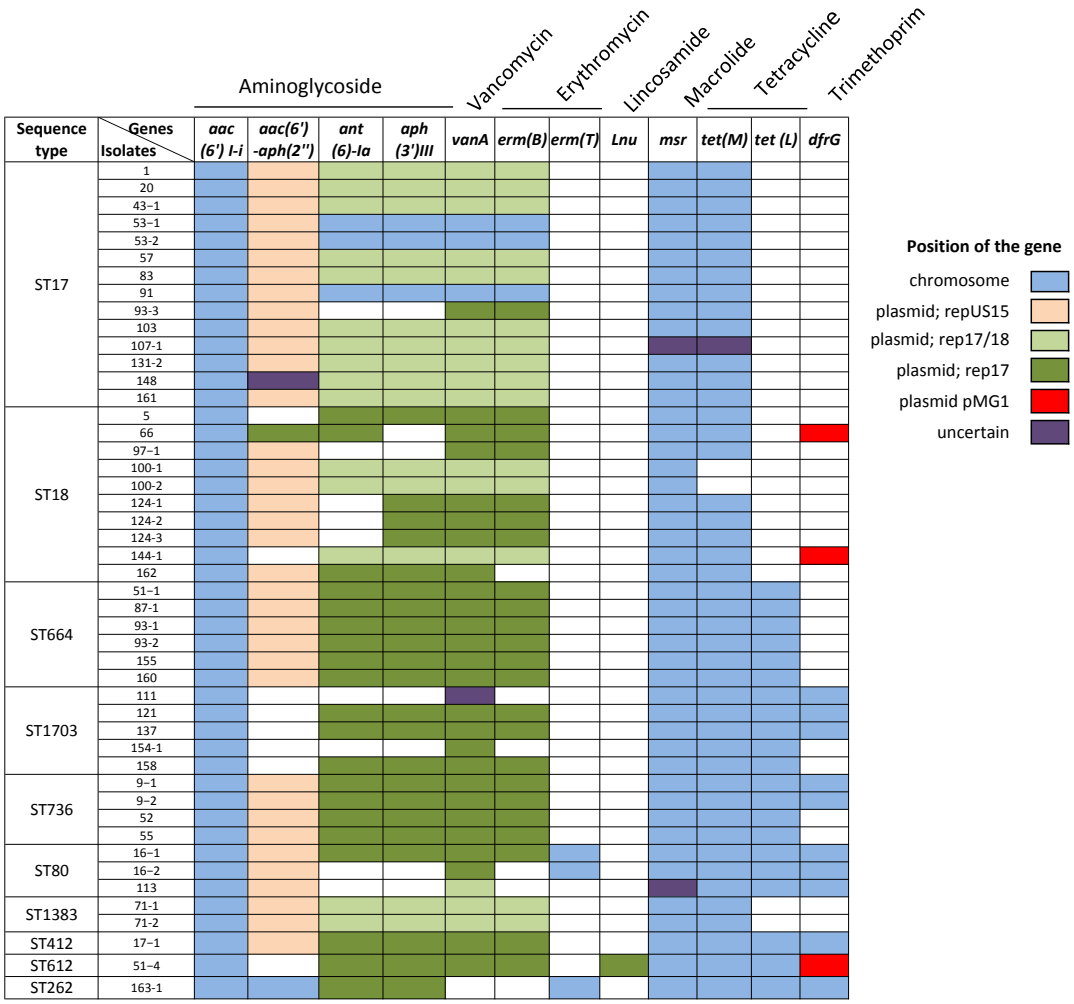


Figure 3. Distribution of antibiotic resistance genes among Dallas isolates. Shown are 12 resistance genes (6'-N-aminoglycoside acetyltransferase, (*aac(6')I-i*), bifunctional enzyme 6'-aminoglycoside acetyltransferase-2"-aminoglycoside phosphotransferase (*aac(6')-aph(2")*), aminoglycoside nucleotidyltransferase (*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).

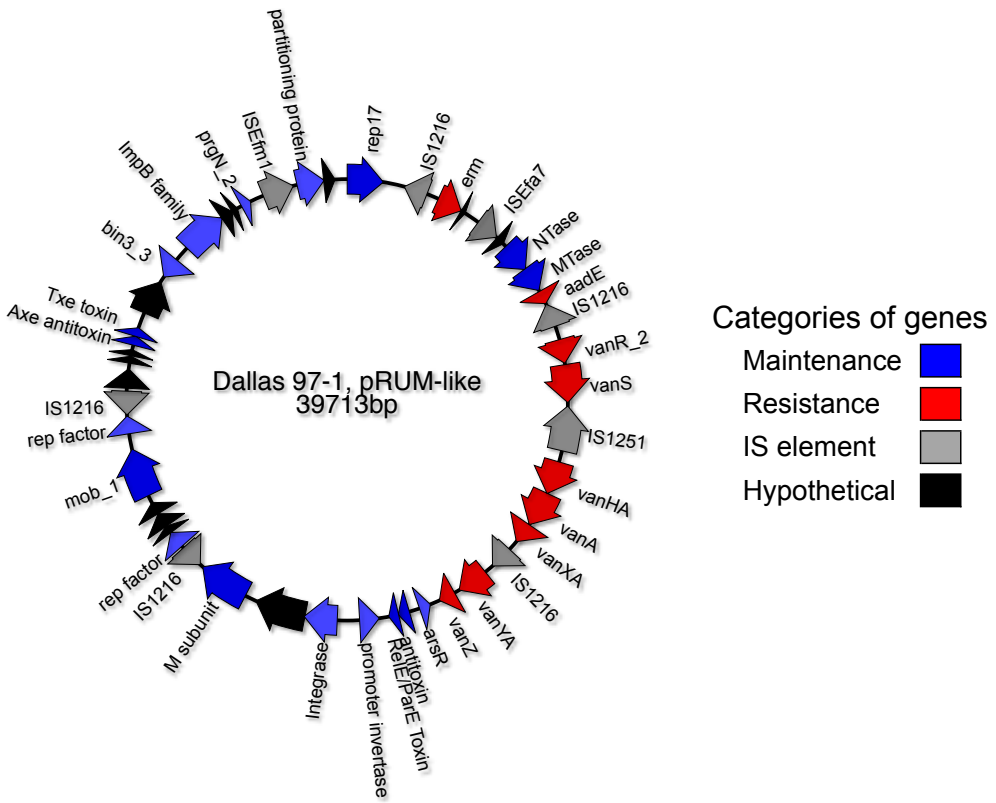


Figure 4. Genetic map of a representative Group 2a pRUM-like plasmid. The 39 kb plasmid p97_1_39kb (accession number CP066571) from the Dallas VREfm isolate 97-1 is shown. Arrows represent genes. The genes are color-coded based on their predicted activity. Maintenance genes are predicted to be responsible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

A.

		Group 1	Group 2a	Group 2b	Group 3	Group 4
rep	*rep18*	+	-	-	-	-
	rep17	+	+	+	+	+
TA	*axe*	+	+	+	-	-
	relE	+	+	+	+	-

B.

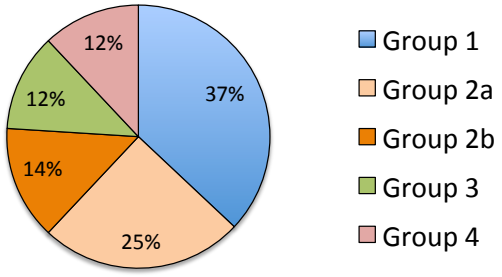
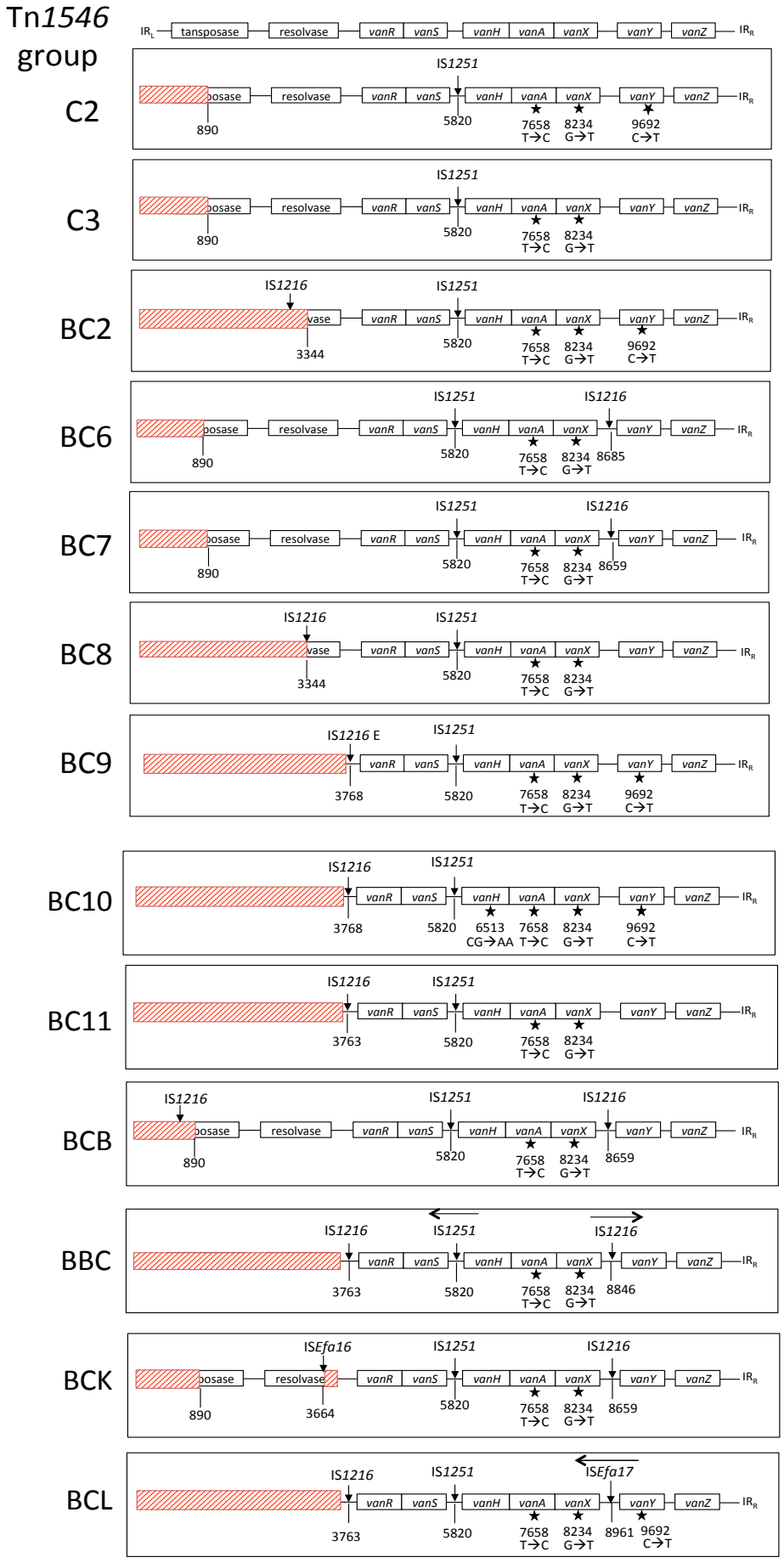


Figure 5. Classification and distribution of pRUM-like plasmid groups. A) pRUM-like plasmids were classified based on the presence or absence (indicated by a + or -, respectively) of 2 replication modules (*rep17* and *rep18*) and 2 stability modules (toxin-antitoxin system; $TA_{axe\text{-}txe}$ and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


Tn1546 group
IR_L
tansposase
resolvase
vanR
vanS
vanH
vanA
vanX
vanY
vanZ
IR_R
C2
IS1251
890
5820
7658 T→C
8234 G→T
9692 C→T
C3
IS1251
890
5820
7658 T→C
8234 G→T
BC2
IS1216
IS1251
3344
5820
7658 T→C
8234 G→T
9692 C→T
BC6
IS1251
IS1216
890
5820
7658 T→C
8234 G→T
8685
BC7
IS1251
IS1216
890
5820
7658 T→C
8234 G→T
8659
BC8
IS1216
IS1251
3344
5820
7658 T→C
8234 G→T
BC9
IS1216 E
IS1251
3768
5820
7658 T→C
8234 G→T
9692 C→T
BC10
IS1216
IS1251
3768
5820
6513 CG→AA
7658 T→C
8234 G→T
9692 C→T
BC11
IS1216
IS1251
3763
5820
7658 T→C
8234 G→T
BCB
IS1216
IS1251
IS1216
890
5820
7658 T→C
8234 G→T
8659
BBC
IS1216
IS1251
IS1216
3763
5820
7658 T→C
8234 G→T
8846
BCK
ISEfa16
IS1251
IS1216
890
3664
5820
7658 T→C
8234 G→T
8659
BCL
IS1216
IS1251
ISEfa17
3763
5820
7658 T→C
8234 G→T
8961
9692 C→T

963

964

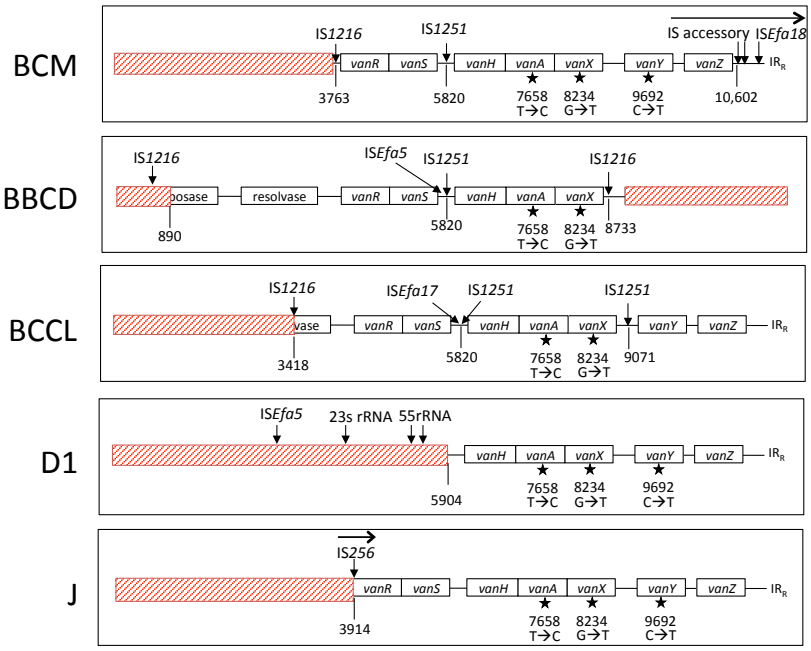


Figure 6: Tn*1546* structural variations in the Dallas VREfm isolates. A reference Tn*1546* structure that includes transposase, resolvase and *van* gene cluster flanked by IR_L and IR_R (inverted repeat left and right, respectively) is delineated at the top. The Tn*1546* groups are represented below the reference Tn*1546*, and the name of the group (C3, BC2, BC6, BC7, BC8, BC9, BC10, BC11, BCB, BBC, BCK, BCL, BCM, BBCD, BCCL D1, J) is shown on the left. Red boxes represent deleted regions relative to the reference Tn*1546*. Vertical lines with numbers at the bottom represent the nucleotide position within the reference Tn*1546.* IS elements are represented by downward arrows. Black stars indicate point mutations. The figure is not drawn to scale.

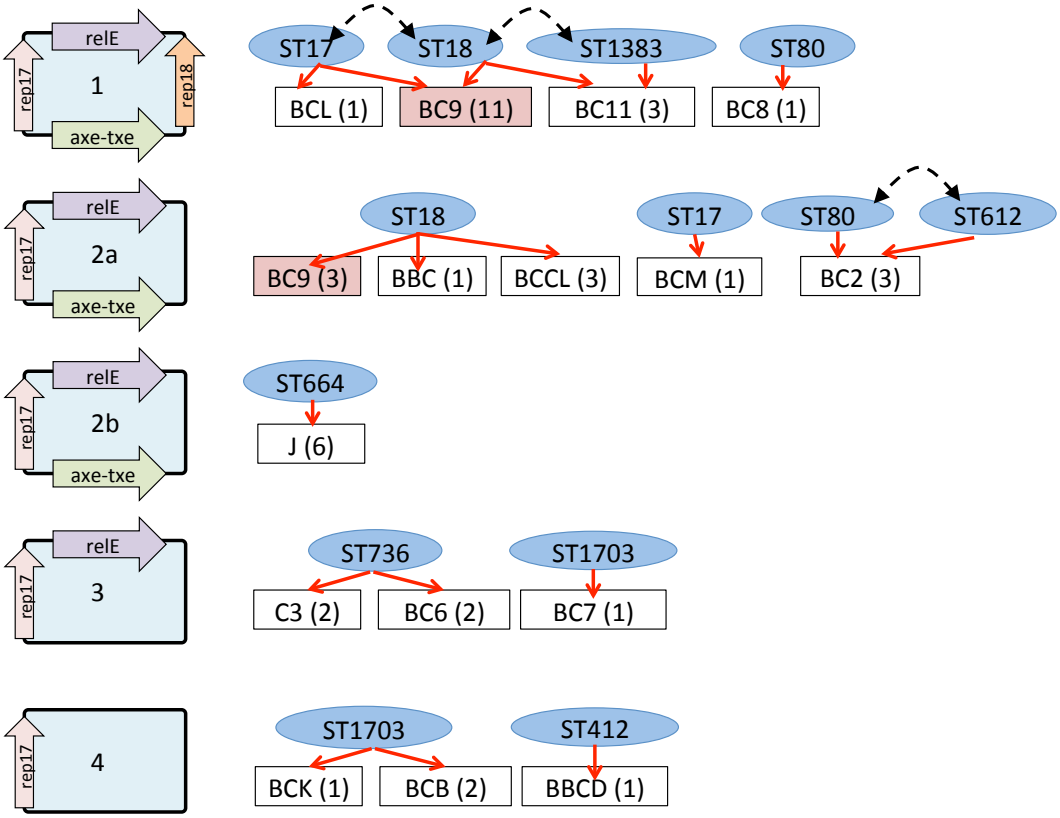


Figure 7. Evolutionary relationships among pRUM-like plasmid groups, Tn*1546*, and VREfm genetic background. The resistance plasmid groups 1, 2a, 2b, 3, and 4 are represented on the left side. The Tn*1546* variations encoded in the corresponding plasmid backbone are displayed on the right side in white boxes. The numbers in parentheses represent total numbers of isolates harboring this Tn*1546* variant on this specific plasmid group belong to this genetic background (ST). The Tn*1546* 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 Tn*1546* variation on the specific plasmid backbone is indicated with blue ovals. A solid red arrow links the ST and the Tn*1546* 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.