Vancomycin resistance in *Enterococcus faecium* from the Dallas, Texas area is conferred predominantly on pRUM-like plasmids

Moutusee Islam,[a] Belle Sharon,[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: Genomic analysis 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

Vancomycin-resistant *E. faecium* (VREfm) is a significant public health concern because of limited treatment options. Genomic surveillance can be used to monitor VREfm transmission and evolution. Genomic analysis of VREfm has not been reported for the Dallas/Fort Worth/Arlington, Texas, area, which is currently the 4th largest metropolitan area in the United States. Our study aimed to address this gap in knowledge by analyzing the genomes of 46 VREfm and one vancomycin-sensitive comparator collected during routine fecal surveillance of high-risk patients upon admission to a Dallas, Texas hospital system (August to October, 2015). 31 complete and 16 draft genome sequences were generated. The closed VREfm genomes possessed up to 12 extrachromosomal elements each. Overall, 251 closed putative plasmid sequences assigned to previously described and newly defined *rep* family types were obtained. Phylogenetic analysis identified 10 different sequence types (STs) among the isolates, with the most prevalent being ST17 and ST18. Strikingly, all but three of the VREfm isolates encoded *vanA*-type vancomycin resistance within Tn*1546*-like elements on a pRUM-like (*rep17*) plasmid backbone. Relative to a previously reported typing scheme for the *vanA*-encoding Tn*1546*, new variants of the Tn*1546* were identified 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*). We conclude that pRUM-like plasmids are important vectors for vancomycin resistance in the Dallas, Texas area and should be a focus of plasmid surveillance efforts.

53 **Importance**
54
55 Vancomycin is an antibiotic used to treat infections caused by multidrug-resistant
56 Gram-positive bacteria. Vancomycin resistance is common in clinical isolates of the
57 Gram-positive pathogen *Enterococcus faecium.* In *E. faecium*, vancomycin resistance
58 genes can be disseminated by plasmids with different host ranges and transfer
59 efficiencies. Surveillance of resistance plasmids is critical to understanding antibiotic
60 resistance transmission. This study analyzed the genome sequences of VREfm
61 collected from the Dallas, Texas area, with particular focus on the mobile elements
62 associated with vancomycin resistance genes. We find that a single plasmid family, the
63 pRUM-like family, is associated with vancomycin resistance in the majority of isolates
64 sampled. Our work suggests that the pRUM-like plasmids should continue to be studied
65 to understand their mechanisms of maintenance, transmission, and evolution in
66 VREfm.
67
68

69 **Introduction**

70 Originally described as a commensals of the healthy human gut, the enterococci
71 attracted attention in the 1990s due to their alarming resistance to multiple antibiotics
72 (1). Vancomycin-resistant enterococci (VRE) were first reported in 1988 in Europe (2)
73 and in the USA in 1989 (3) and are a major concern (4). Vancomycin-resistant *E.*
74 *faecium* (VREfm) is particularly concerning due to limited treatment options and
75 widespread occurrence of vancomycin resistance among infection isolates of *E.*
76 *faecium* (4).

77

78 Phylogenetic analysis of *E. faecium* strains revealed the presence of two major clades,
79 referred to as Clade A and B, with some evidence for subclades A1 and A2 within Clade
80 A (5, 6). Clade B *E. faecium* strains are mostly commensals of the healthy human gut.
81 The multidrug-resistant *E. faecium* strains responsible for hospital outbreaks typically
82 belong to Clade A1, whereas animal gut commensals typically belong to Clade A2 (6).
83 Multi-locus sequence typing (MLST) pre-dates whole genome-based phylogenetic
84 analyses and has been used globally to classify *E. faecium* isolates by nucleotide
85 sequence variations occurring in 7 housekeeping genes (7). Isolates are assigned to
86 different sequence types (ST) based on the allelic variants. Most reported hospital
87 outbreaks of VREfm have emerged from a single genetic lineage, referred to as clonal
88 complex-17 (CC17), that was founded by the sequence type ST17 (8).

89

90 Vancomycin resistance genes are encoded within mobile elements in VREfm and can
91 be horizontally disseminated (9-11). VanA-type resistance commonly occurs among
92 VREfm, and VanA-type resistance genes are typically encoded within Tn*1546* (12).
93 Tn*1546* may be chromosomally integrated or plasmid-borne. Different plasmid
94 backbones carrying Tn*1546*, including Inc18, pRUM, and pLG1, have been reported

95 worldwide (for *e.g.* references (10, 13)). Structural variations of Tn*1546* due to the
96 presence of insertion sequences (IS) including IS*1251,* IS*1216,* IS*1485* and IS*Ef1* have
97 also been reported. These previous studies characterized Tn*1546* structural variation
98 by PCR mapping and then sequencing of the overlapping fragments (10, 13, 14).
99 However, in the absence of a completely closed genome assembly, conclusive links
100 between *vanA*, Tn*1546* variants, and specific plasmid backbones in VREfm can be
101 difficult to achieve.
102
103 Surveillance of antibiotic resistance elements is important because changes in their
104 host range and transfer frequency could evolve, which could impact clinical care.
105 Utilization of whole genome sequencing for epidemiologic study provides an
106 opportunity for multilevel analysis that includes detection of novel mobile genetic
107 elements and monitoring DNA sequence changes in resistance gene vectors. The
108 advent of long-read next generation sequencing (NGS) by Pacific Biosciences and
109 Oxford Nanopore Technology (ONT) has made it possible to generate genome
110 assemblies of a large number of bacterial isolates in a cost-effective manner. The
111 combination of long (ONT in this study) and short (typically Illumina) reads generates
112 high quality, completely closed genome assemblies with fully assembled mobile genetic
113 elements (15).
114
115 In this study, we analyzed phylogenetic relationships and plasmid diversity among a
116 previously reported collection of VREfm from the Dallas, Texas area (16). These
117 isolates were recovered from rectal surveillance of high-risk patients upon admission
118 to a Dallas hospital system. These represent colonization reservoirs from which future
119 hospital outbreaks could emerge. Some results of our study are expected, including the
120 predominance of CC17/Clade A1 among the VREfm isolates. Other results highlight

121 important areas for future study. Specifically, we note the high number of
122 extrachromosomal elements in VREfm, with some isolates possessing 12 unique
123 elements independent from the chromosome. How these multiple elements are
124 coordinated and maintained in VREfm, and what they contribute to VREfm physiology,
125 are largely unknown and are important topics for future investigation. Moreover, pRUM-
126 like plasmids are the dominant carriers of vancomycin resistance genes among these
127 VREfm. This plasmid family should continue to be investigated in laboratory
128 experiments to understand its maintenance and transmission mechanisms.
129
130 **Materials and Methods**
131
132 **Bacterial strains analyzed in this study**. The isolates used in this study (**Dataset**
133 **S1A**) were cultured from surveillance rectal swabs of high-risk patients on hospital
134 admission (August 2015 to October 2015). Their clinical collection and initial
135 characterization was previously reported (16) and is briefly summarized here. Rectal
136 swabs from patients admitted with at least one of the following risk factors
137 (hospitalization for ≥2 consecutive nights in the preceding 30 days, transferred from
138 another medical facility, residence in a nursing home or extended/long term care facility,
139 or the presence of decubitus ulcer or a draining wound) were cultured on Spectra VRE
140 (Remel) agar as a part of routine infection control surveillance. Instead of disposal after
141 clinical procedures were complete, cultures positive for presumptive VRE growth were
142 coded numerically, de-identified, and transferred to the University of Texas at Dallas
143 for further analysis. Isolates were coded by de-identified patient number. If multiple
144 colonies were collected from a Spectra VRE plate, this was indicated by hyphenated
145 numerics. For example, the VRE isolates 124-1, 124-2 and 124-3 were collected from

146 the same rectal swab sample from the same patient (de-identified patient #124). No
147 patients were sampled longitudinally.
148
149 In our previous study of these isolates, presumptive VREfm were confirmed by *ddl*
150 typing and for the presence of *vanA* or *vanB* genes by PCR (16, 17). All isolates used
151 in this study, with the exception of 163-1, initially grew on brain heart infusion (BHI)
152 agar supplemented with 256 μg/mL vancomycin. The vancomycin MIC of isolate 163-
153 1 was previously determined to be 2 μg/mL by broth microdilution (16). The vancomycin
154 MIC of 10 additional isolates were previously determined to be >256 μg/mL by broth
155 microdilution (16). The vancomycin MIC of all other isolates were determined in this
156 study using the same broth microdilution method previously reported (16).
157
158 **Genome sequencing and assembly.** Methods for DNA isolation, MinION sequencing,
159 Illumina sequencing, and hybrid genome assembly are described in the Supplemental
160 Text.
161
162 **Phylogenetic analyses and MLST**. Genome assemblies were annotated with
163 PROKKA (v 1.12) (18). A sequence alignment was generated for 1706 core genes by
164 ROARY (v 3.12) (19) using PRANK where the paralogs were not split. A phylogenetic
165 tree was built by RAxML (v 8.2.12) (20) with rapid bootstrapping (1000 inferences)
166 using the GTR+GAMMA model and subsequent ML search on the core gene alignment
167 where the outgroup was specified to be the reference genome *E. faecium* 1,231,501
168 (GenBank accession number ACAY01). The phylogenetic tree was visualized with
169 iTOL (v4) (21). MLST was determined by the *E. faecium* MLST database
170 (https://pubmlst.org/efaecium/) (7, 22). The novel sequence type ST1703 and novel *gyd*

171 allele number (allele-70) was assigned for the 5 VREfm isolates 111, 121, 137, 154-1,
172 158 by request from the database.
173
174 **Plasmid detection and typing**. Each contig of <2 Mbp in size in a closed genome
175 assembly was considered a putative plasmid. *In silico* detection and typing of plasmids
176 was performed with PlasmidFinder (v 2.1) (23). Putative plasmids in closed genome
177 assemblies that were not classifiable by PlasmidFinder were analyzed further (see
178 below).
179
180 For comparison with pRUM-like/*rep17* plasmids from Dallas isolates, plasmids were
181 extracted from PLSDB v.2021_06_23_v2 (24, 25) with the following terms: genus =
182 *Enterococcus*, topology = circular, PlasmidFinder = contains *rep17*. Plasmids were
183 visualized with Easyfig (26).
184
185 **Definition of pMIX plasmid groups**. Putative plasmid sequences that were
186 unclassifiable by PlasmidFinder were analyzed as follows. Presumptive *rep* genes were
187 identified from plasmid annotations and analyzed by NCBI Conserved Domain Analysis
188 (CDS). Confirmed *rep* genes were those that encoded protein domains that matched
189 known Rep protein families with e-values of e^{-10} to e^{-78}. The Rep sequences were also
190 scanned by HMMER (v 2.41.1) (27) to analyze Pfam domains. Multiple sequence
191 alignment of the nucleotide and predicted amino acid sequences of *rep* genes were
192 performed with MUSCLE (v 3.8.424) where the “group sequences by similarity” option
193 was selected. A neighbor-joining tree was built based on the sequence alignment using
194 the “Tamura Nei” genetic distance model (28). The tree was visually scrutinized in order
195 to sort the plasmids into 9 groups harboring *rep* genes with at least 95% identity in both
196 nucleotide and predicted amino acid sequence (except for pMI4; see Results). Each

197 unique *rep* group of plasmids was given a name of format pMIX, where the "X" is a
198 numeric number from 1 to 9. A representative sequence from each pMIX *rep* group is
199 in **Dataset S1B**. Representative sequences were compared to PLSDB
200 v.2021_06_23_v2 (24, 25) using BLASTn with 90% identity and query coverage
201 thresholds.
202
203 Eight closed circular DNA elements of varying sizes (3 to 72 kb) could not be classified
204 by PlasmidFinder nor by the pMIX *rep* typing scheme described above. Analysis of
205 these elements is described in the Supplemental Text.
206
207 **Average nucleotide identity (ANI) analysis.** All vs. all ANI was calculated using
208 ANIclustermap v1.1.0 at default parameters with either all *E. faecium* assemblies or
209 *rep17*/pRUM plasmids as input.
210
211 **Transcriptional activity of predicted toxin-antitoxin (TA) systems**. Transcriptional
212 activity of the TA systems, $TA_{axe\text{-}txe}$ and TA_{relE}, was determined by RT-qPCR for the
213 VREfm isolates 1 and 5, described in the Supplemental Text. The primer sequences
214 used for qPCR are provided in **Dataset S1C.**
215
216 **Resistance gene identification**. Acquired antimicrobial resistance genes were
217 detected with ResFinder 3.1 (29).
218
219 **Tn*1546* analysis.** The Tn*1546* nucleotide sequence described in *E. faecium* BM4147
220 (accession number M97297) was used as a reference to identify variations occurring
221 in the Tn*1546* identified in this study. The Tn*1546* were classified using a previously
222 described nomenclature (13). The presence of insertion sequence elements (IS

223 elements) was denoted with a one-letter code, shown in parentheses: IS*1216* (B),
224 IS*1251* (C), IS*Efa5* (D), IS*256* (J), IS*Efa16* (K), IS*Efa17* (L), IS*Efa18* (M). The
225 combination of IS elements within the transposon is described by a two or three letter
226 code; i.e., group 'BC' possesses one each of IS*1216* and IS*1251*. Arabic numerals,
227 following the alphabet code, indicate differences due to point mutations or different IS
228 element insertion sites. For the group 'BC', a previous study (13) described BC1-BC5,
229 each with specific insertion sites of IS elements. Any novel insertion sites identified in
230 this present study were therefore numbered from 'BC6' forward. The novel IS*Efa16,*
231 IS*Efa17,* IS*Efa18* sequences were identified in this study and registered in the ISFinder
232 public database (30).
233
234 **Accession numbers.** DNA sequences from this study have been deposited under
235 Bioproject PRJNA682584. Individual accession numbers for all sequence files are
236 provided in **Dataset S1A**.
237
238 **Results**
239 **ST17 and ST18 VREfm isolates were prevalent in the collection.** Genome
240 sequencing and hybrid assembly were performed for 47 *E. faecium* isolates with Oxford
241 Nanopore MinION and Illumina technologies. 31 closed and 16 draft genome
242 assemblies were generated (**Dataset S1A**). In our previous analysis of these isolates
243 using PCR, all isolates were found to be *vanA*-positive and *vanB*-negative, with the
244 exception of the isolate 163-1, which had neither *vanA* or *vanB* and had a vancomycin
245 MIC of 2 μg/mL by broth microdilution (16). We included 163-1 in this study for
246 comparative purposes to evaluate its relationship to the *vanA*-positive isolates in the
247 collection. After recovery from Spectra VRE rectal surveillance cultures, all 46 *vanA*-
248 positive isolates sequenced in this study grew on agar supplemented with 256 μg/mL

249 vancomycin. Subsequent broth microdilution assays performed over a year later after
250 genome sequencing and freezer restocking re-confirmed vancomycin resistance for all
251 isolates except for 9-2 and 154-1, which had vancomycin MICs ≤ 4 μg/mL. Likely, loss
252 of vancomycin resistance elements during laboratory culture led to these phenotypes;
253 however, this was not investigated further in this study.
254
255 Phylogenetic analysis based on alignment of 1706 core genes allowed us to assess
256 the diversity of the VREfm isolates (**Figure 1**). Two reference genomes, *E. faecium*
257 1,231,501 (5) and *E. faecium* ATCC® 700221™(31), were included as representatives
258 of *E. faecium* clade A2 and A1, respectively. *E. faecium* 1,231,501 is a human
259 bloodstream isolate from the USA and is susceptible to vancomycin. *E. faecium*
260 ATCC® 700221™ is a USA isolate and VanA-type VREfm control strain used in
261 antimicrobial susceptibility testing. As expected, the Dallas VREfm isolates clustered
262 with the clade A1 reference *E. faecium* ATCC® 700221™.
263
264 Ten different sequence types (ST), all belonging to clonal complex 17 (CC17), were
265 identified in the collection (**Figure 1, Figure 2**). Isolates belonging to ST17 and ST18
266 are the most prevalent in the collection, comprising 50% of the isolates, which clustered
267 together in the core genome tree (**Figure 1, Figure 2**). Six STs (ST18, ST664, ST1703,
268 ST736, ST80 and ST612) are double locus variants of ST17. ST262 and ST412 are
269 triple locus variants and ST1383 is a single locus variant of ST17 (**Dataset S1D**).
270 ST1703, a double locus variant of ST17 (*pstS* allele-20*;* novel *gyd* allele-70), is a novel
271 ST in the MLST database and comprised 11% of the collection. We conclude that *E.*
272 *faecium* of multiple different STs, all within CC17, colonized the Dallas patients sampled
273 at the time of the study.
274

275 We performed an additional analysis to determine how closely related *E. faecium*
276 strains are in our collection. All vs. all Average Nucleotide Identity (ANI) was calculated
277 with all *E. faecium* assemblies as input (**Dataset S1E**). Dallas *E. faecium* isolates
278 possess ≥98.98% ANI, consistent with the core gene phylogeny in **Figure 1**. In most
279 cases, strains isolated from the same patient were identical or nearly so. For example,
280 the three isolates from patient 124 shared 99.99% ANI. Similar observations were made
281 for isolates from patients 100, 16, 53, 71, and 9. Exceptions were patients 51 and 93,
282 which each were colonized with different *E. faecium* STs (**Figure 1**), also reflected in
283 the ANI analysis of the isolates (≤99.4% ANI).
284
285 **Large numbers of plasmids are present in the Dallas isolates**. The plasmid content
286 of the Dallas isolates was analyzed by an *in silico* detection method utilizing the
287 conserved nucleotide sequences of replication initiation (*rep*) genes (23). Each contig
288 of <2 Mbp in size in a closed genome assembly was considered a putative plasmid. Of
289 the 259 closed circular putative plasmids, 141 were assigned to 10 known *rep* families
290 described in the PlasmidFinder database. The remaining 118 putative plasmids were
291 categorized into 9 novel plasmid groups (pMI1 to pMI9) and 8 other elements
292 (discussed in the next section).
293
294 The Dallas isolates harbored a variable number of putative plasmids (n=3 to 12) of
295 sizes ranging from 1.9 to 294 kb (**Figure 1; Dataset S1F**). The most common elements
296 were a megaplasmid (126 to 294 kb) and a pRUM-like plasmid (17 to 63 kb), which
297 were each present in all isolates except 144-1, which lacked the megaplasmid. Both
298 the reference strains (*E. faecium* 1,231,501 and ATCC® 700221™) harbored the
299 megaplasmid, but 1,231,501 lacked the pRUM-like plasmid. Each of the megaplasmids
300 harbored the *rep* gene *repUS15*, encoding a RepA_N protein motif. The pRUM-like

301 plasmid harbored *rep17* or both *rep17* and *rep18*, which we refer to as a mosaic
302 plasmid. An analysis of plasmid replicon enrichment among specific STs is presented
303 in the **Supplemental Text** and **Dataset S1G**.
304
305 **Uncategorized plasmids were sorted into 9 plasmid groups, pMI1 to pMI9.**
306 PlasmidFinder (23) did not identify *rep* genes in a total of 118 circular entities. Of these
307 118 circular entities, 110 were established to be plasmids based on the presence of
308 *rep* genes not currently represented in the PlasmidFinder database (see Materials and
309 Methods). The remaining 8 circular entities were designated as putative excision
310 products of genomic islands (**Supplementary Text, Dataset S1H-I**) including a 5 kb
311 circular fragment of Tn*1546* in isolate 111.
312
313 The 110 plasmids not classifiable by PlasmidFinder were sorted into 9 plasmid groups,
314 pMI1 to pMI9, based on phylogenetic analyses of the *rep* genes (**Figure S1; Dataset**
315 **S1J**). The nucleotide sequences of the *rep* genes and their corresponding amino acid
316 sequences shared >95% identity within each pMIX plasmid group, except for pMI4
317 (**Table 1**). The clustering of the *rep* gene sequences belonging to the pMI4 plasmid
318 group supported their inclusion in a single group, despite lower pairwise sequence
319 identities (**Figure S1**). A total of 6 Rep protein families (RepA_N, Rep_3, Rep_1,
320 Rep_trans, Rol_rep_N, and Rep_2) were encoded by the *rep* genes of the 9 pMIX
321 groups (**Dataset S1J**). Specifically, the *rep* genes of pMI1 and pMI3 encoded protein
322 families RepA_N and Rep_1, respectively. The *rep* genes of pMI2, pMI5, pMI6 and
323 pMI7 encoded the protein family Rep_3. The *rep* genes of pMI4 and pMI8 encoded
324 protein families Rep_trans and Rep_2, respectively. The *rep* of pMI9 was the only
325 exception, encoding two Rep protein families, Rep_trans and Rol_rep_N (**Table 1**).
326

327 Each of the Dallas isolates harbored 1 to 5 of the pMIX plasmid groups, except the
328 isolate 83 (ST17), which had none of them (**Figure 1**). Neither of the two reference
329 strains harbored any of the pMIX plasmid groups. The pMI1 was the most frequently
330 occurring pMIX plasmid group, identified in 75% of the isolates (**Table 1**). The pMI8
331 and pMI9 were the rarest ones, found in only isolates 51_1 (ST612) and 5 (ST18),
332 respectively. The pMIX plasmid groups were not enriched in any specific ST. We
333 observed that plasmids within a very narrow size range (standard deviation 0.2 to 2.3)
334 were grouped together by this *rep* typing scheme (with the exception of pMI1), although
335 plasmid size was not taken into consideration in assigning plasmid groups.
336
337 We used PLSDB (24, 25) to compare representative *rep* sequences from the pMIX
338 plasmid groups with existing plasmids in the database (as of January 2023). PLSDB
339 compiles plasmid sequences and their associated metadata, including PlasmidFinder
340 and other analyses. This was done as an additional check on our analysis, as ours
341 was initially performed in 2018-2019. We reasoned that the pMIX typing scheme
342 would be most useful if it encompassed other previously untypeable plasmids in the
343 database. We applied strict thresholds of 90% identity and 90% query coverage to
344 filter BLASTn hits to the database. The results support pMI1, pMI3, pMI5, pMI6, and
345 pMI8 as newly defined *rep* types that include other previously untyped *Enterococcus*
346 plasmids from around the world (**Dataset S1K**). The only hit to pMI9 *rep* was the
347 Dallas isolate from which it was defined. In this PLSDB analysis, pMI2 was assigned
348 to *rep11a*, and pMI4 was assigned to *rep14b*. The pMI7 analysis was more
349 complicated, with one group of identical hits (100% identity and query coverage) for
350 many plasmids with no known *rep* type assigned, and another group of hits with 100%
351 identity but lower 98% coverage that were assigned to *rep17*.
352

353 **Resistance genes were encoded in both chromosomes and plasmids**. Different
354 combinations of 12 resistance genes predicted to provide resistance against 5 classes
355 of antibiotics (aminoglycosides, glycopeptides, macrolides, tetracycline, and
356 trimethoprim) were identified among the Dallas isolates (**Figure 3**; **Dataset S1L;**
357 **Supplemental Text**). The most commonly occurring resistance genes were *aac(6')I-1*
358 (100% of isolates), *msr* (100%), *vanA* (95%), *tetM* (95%), *ermB* (85%), and *aph(3')*
359 (85%). These resistance genes were encoded in the chromosome and primarily in only
360 two plasmid types, the pRUM-like plasmid group and the megaplasmid. The pRUM-like
361 plasmids predominantly encoded 4 of the resistance genes (*ant(6)-Ia, aph(3')III, vanA,*
362 and *ermB*) conferring resistance against aminoglycosides, vancomycin and
363 erythromycin (**Figure 3**). The *repUS7* plasmid (pMG1) was found among only 3 of the
364 Dallas isolates - 66 (ST18), 144-1 (ST18) and 51-4 (ST612) - and encoded *dfrG*, for
365 trimethoprim resistance. We note that antibiotic resistance of these isolates was not
366 phenotypically assessed in this study, other than for vancomycin resistance.
367
368 All of the Dallas isolates, except for 163-1, encoded VanA-type vancomycin resistance
369 genes. The *vanA* genes possessed 99.9% nucleotide and 99.7% amino acid sequence
370 identity with the reference *vanA* sequence (GenBank AAA65956.1 (**Dataset S1L).** The
371 reference sequence is encoded within the transposon Tn*1546* (GenBank M97297),
372 originally described in the human fecal isolate *E. faecium* BM4147 collected from
373 France in 1988 (32) (**Figure S2**). A classical structure of the *vanA* gene cluster,
374 consisting of *vanHAX* and *vanYZ* encoded within Tn*1546* elements, was observed
375 among the Dallas VREfm isolates, except for the isolate 17-1, which lacked the genes
376 *vanY* and *vanZ*. The Tn*1546* structure varied among the isolates in terms of insertion
377 site (chromosome versus plasmid), IS elements, and point mutations, discussed further
378 below.

379

380 **The Tn*1546* is borne on pRUM-like plasmids in most isolates**. The complete
381 Tn*1546* was carried by the pRUM-like plasmid in 42 of 46 *vanA*-positive isolates. A
382 representative pRUM-like plasmid from our collection is shown in **Figure 4**. The isolate
383 111 did not encode the entire Tn*1546* in the pRUM-like plasmid; rather, it encoded
384 *vanRS* in a 17 kb pRUM-like plasmid while *vanHAX* was encoded within a separate 5
385 kb circular element (te111_5kb) (**Dataset S1H**). Tn*1546* was chromosomally encoded
386 in 3 isolates, 53-1, 53-2 and 91, each belonging to ST17, with all three strains also
387 harboring a pRUM-like plasmid lacking Tn*1546* (**Table 2**).

388

389 We categorized the 43 pRUM-like plasmids encoding Tn*1546* or a fragment of Tn*1546*
390 (for isolate 111) into four groups based on the presence or absence of replication and
391 stability modules on their backbones (**Table 2, Figure 5**), per a previously described
392 typing scheme (14). For the plasmids in our collection, two toxin-antitoxin (TA) systems
393 were observed in different distributions, *axe-txe* ($TA_{axe\text{-}txe}$) and a *relE/parE* family toxin
394 associated with an uncharacterized antitoxin component (TA_{relE}) (**Table 2**). Both of the
395 TA systems were actively expressed in two representative VREfm isolates (**Figure S3**).

396 i) Group 1 ($rep_{17;}$ rep_{18}; $TA_{axe\text{-}txe;}$ TA_{relE}; n=16): These plasmids possess $rep_{17/18}$
397 and both $TA_{axe\text{-}txe}$ and TA_{relE}. All the group 1 plasmids encoded the *rep18b*
398 (accession number CP018068), except isolate 113, which encoded the
399 *rep18a* (accession number AB158402). The plasmids had a narrow size
400 range (44 to 48 kb) and were present in VREfm isolates from ST17, ST18,
401 ST1383 and ST80.

402 ii) Group 2 (rep_{17}; $TA_{axe\text{-}txe}$; TA_{relE}; n=17): These plasmids possess rep_{17} and
403 both $TA_{axe\text{-}txe}$ and TA_{relE}. The group 2 plasmids were classified further based
404 on size and genetic background of the isolates carrying them. Group 2a

405 plasmids (29-45 kb) were carried by the reference strain ATCC® 700221™,
406 and isolates belonging to ST18 (5, 66, 97-1, 124-1, 124-2, 124-3, and 162),
407 ST80 (16-1, 16-2), ST17 (93-2), and ST612 (51-4). The comparatively larger
408 (57-63 kb) group 2b plasmids were present in all 6 isolates belonging to
409 ST664 (51-1, 87-1, 93-1, 93-2, 155, and 160). Gene presence/absence
410 analysis using Roary (19) identified a set of genes that are specific to Group
411 2b plasmids (**Dataset S1M**), including a ~7.5 kb region encoding helix-turn-
412 helix XRE-family like proteins (cd00093) inserted immediately upstream of
413 the *van* gene cluster (**Figure S4**).

414 iii) Group 3 (rep_{17}; TA_{relE}; n=5): These plasmids possess rep_{17} and TA_{relE}, and
415 have a size range of 36-42 kb. They were carried by ST736 isolates (isolates
416 9-1, 9-2, 52, and 55) and one ST1703 isolate (isolate 121).

417 iv) Group 4 (rep_{17}; n=5): These plasmids possess rep_{17} and no known stability
418 modules, and have a size range of 17-35 kb. These plasmids were carried
419 by isolates belonging to the novel genetic background ST1703 (111, 137,
420 154-1, and 158) and by a single ST412 isolate (17-1).

421

422 We further analyzed the pRUM-like plasmids in our collection using two approaches: a
423 core genome phylogeny based on sequence variation in 7 pRUM-like plasmid core
424 genes (*rep17,* IS*1216, vanS,* ImpB/MucB/SamB family protein, hypothetical protein,
425 plasmid partitioning protein, and replication control protein PrgN) (**Figure S5**), and an
426 analysis of ANI **(Figure S6; Dataset S1N).** The pRUM-like plasmids form two clusters
427 in the phylogenetic tree (**Figure S5)**, which correspond to the presence/absence of
428 $TA_{axe\text{-}txe}$. The pRUM-like plasmids ranged from 93.9 – 100% pairwise ANI. ANI analysis
429 identified well-defined clusters of plasmids with high percent ANI that correspond to
430 plasmid groups 2b, 3, and 4, while groups 1 and 2a are intermixed (**Figure S6**). This is

431 also consistent with the presence/absence of $TA_{axe\text{-}txe}$ being an important discriminator
432 for grouping of pRUM-like plasmids.
433
434 Finally, we compared pRUM-like/*rep17* plasmids from the Dallas collection with other
435 *rep17* plasmids that have previously been published. PLSDB was queried to obtain
436 circular plasmids assigned to genus *Enterococcus* and to *rep17*. A strict cut-off of
437 ≥99.9% pairwise ANI was applied, which identified a cluster of closely related pRUM-
438 like/*rep17* plasmids from different sources (**Figure 6**). These plasmids have ≥99.9%
439 ANI, are from geographically distinct regions (USA [this study], France (33), India (34),
440 and Australia (35, 36)), and are present in different *E. faecium* STs. The most closely
441 related are the pRUM-like/*rep17* plasmid from Dallas isolate 16-1 and the *rep17*
442 plasmid from the VREfm strain 15-307.1. These plasmids have identical gene content
443 and differ by SNPs and deletions/insertions at 10 sites. VREfm 15-307.1 was isolated
444 in France in 2015 from a rectal swab of a hospitalized patient, which is similar to the
445 collection of Dallas 16-1 in 2015.
446
447 **Dallas VREfm harbored 17 variants of Tn*1546*.** The Tn*1546* region of the Dallas
448 isolates varied from the typical structure, which consists of ORF1 (transposase), ORF2
449 (resolvase) and the *van* genes flanked by IR_L (inverted repeat, left) and IR_R (inverted
450 repeat, right) (**Figure 7**). A total of 17 different sequence variations in the Tn*1546* were
451 observed among 45 of the Dallas isolates (**Figure 7, Table 3**). These 17 variations
452 were classified using a nomenclature system developed previously (13) based on the
453 presence of IS elements, point mutations, and deletions relative to the reference
454 Tn*1546*. VREfm belonging to the same Tn*1546* group shared 99.8%-100% nucleotide
455 sequence identity in the Tn*1546* region. The isolate 111 was excluded from this
456 grouping since it harbored a fragmented Tn*1546* occurring in a pRUM-like plasmid and

457 another circular entity (**Dataset S1H**). The Tn*1546* variation was mostly associated with
458 the distribution of 7 different IS elements (IS*1251*, IS*1216*, ISEfa*16*, IS*Efa5*, IS*Efa17*,
459 IS*256*, IS*Efa18*), three of which were novel to the Dallas isolates. Most of the Tn*1546*
460 groups described in this study are novel relative to those previously reported in the
461 literature, except for group BC2, which was reported previously in Poland (13). The
462 most predominant Tn*1546* groups were BC9 (n=14) and J (n=6). The BC9 variety was
463 observed in ST17 (n=9) and ST18 (n=5) isolates. The reference strain
464 ATCC® 700221™ harbored the C2 structural variant of Tn*1546* that was described
465 previously (13) but is not present in the Dallas VREfm isolates. The presence of
466 different pRUM groups and Tn*1546* types across the phylogeny of *E. faecium* isolates
467 in this study is presented in **Figure S7**.
468
469 Specific structural variations of the Dallas Tn*1546* elements are described further in the
470 Supplemental Text.
471
472 **Potential VREfm patient-patient transmission.** We used our genome data to look for
473 identical *E. faecium* clones colonizing different patients. This would be suggestive of
474 prior transmission of VREfm between patients in our area. We found one possible
475 instance supported by our genomic data. *E. faecium* isolates from patients 55 and 52
476 were of identical ST, same pRUM plasmid size and group, same Tn*1546* group (and
477 are the only members of the group in this study), and possess 99.99% ANI overall.
478 Their pRUM plasmids possess >99.99% ANI, differing by one SNP. However, the total
479 genome size for these two strains differs. The isolate from patient 55 has a closed
480 genome that is 1026 bp longer than the draft genome of the patient 52 isolate.
481
482

483 **Discussion**

484 This study provided a snapshot of fecal VREfm colonizing hospitalized patients in
485 Dallas, Texas from August – October 2015. VREfm isolates from previously described
486 and new STs were detected. The most prevalent VREfm isolates, comprising 50% of
487 those analyzed, belonged to ST17 and ST18. These STs were previously reported to
488 be responsible for outbreaks in countries including Portugal (14), Spain (37), Ireland
489 (38), Denmark (39), Columbia (40), Brazil (41), Canada (42). The other 8 STs identified
490 among our collection (ST1383, ST612, ST412, ST80, ST1703 and ST664), were single,
491 double, or triple locus variants of ST17. ST1703 is a novel sequence type described in
492 this study. ST182, which caused an outbreak in San Antonio, Texas in the 1990s (43),
493 was not detected in our collection.

494

495 All the plasmid-borne Tn*1546* in Dallas isolates were encoded within pRUM-like
496 plasmids that were clustered into 4 subgroups (1, 2a, 2b, 3, and 4), defined by the
497 presence of the replication module *rep17* and a combination of other replication and
498 stability modules. The patients from whom the isolates in this study were collected
499 resided in different areas of Dallas and in unconnected facilities including nursing
500 homes and personal residences. Therefore, the presence of only one type of plasmid
501 backbone carrying *vanA* among these patients was striking. Our data are consistent
502 with another recent analysis of VREfm from the United States. Chilambi, et al (2020)
503 (44) analyzed gastrointestinal and blood VREfm isolated from pediatric patients at St.
504 Jude's Children's Research Hospital over a 10 year period. VanA-type VREfm was
505 isolated from 23 of 24 patients analyzed, and for all 23, the Tn*1546* was encoded on a
506 *rep17* element. Clearly, these plasmids are significant vectors for vancomycin
507 resistance in *E. faecium,* as proposed by Freitas, et al. (10) in a multilevel analysis of
508 VREfm. In our study, multiple predicted antibiotic resistance genes were detected on

509 these plasmids, in addition to vancomycin resistance genes. The pRUM-like plasmids
510 should be a focus of plasmid surveillance. Resistance genes may be consolidating on
511 this specific backbone, as opposed to the many other plasmids detected in VREfm in
512 this study.
513
514 We detected carriage of up to 12 extrachromosomal elements in the VREfm isolates in
515 our study. This is consistent with the comprehensive work of Arredondo-Alonso, et al.,
516 who conducted a large-scale analysis of the *E. faecium* plasmidome (45). The authors
517 determined that *E. faecium* from hospitalized patients possess the largest plasmidome
518 in terms of plasmid number and total bp. The striking number of extrachromosomal
519 elements among VREfm warrants further investigation into their maintenance,
520 transmission, and functions.
521

522 **Acknowledgements**

523 This project was supported by R01AI116610 and the Cecil H. and Ida Green Chair in
524 Systems Biology Science to K.P. We gratefully acknowledge Dr. Joslyn Pribble and the
525 Microbiology lab at Methodist Health System, without whom this work would not be
526 possible.
527

528 **References**

529 1. Bonten MJ, Willems R, Weinstein RA. 2001. Vancomycin-resistant
530 enterococci: why are they here, and where do they come from? Lancet Infect
531 Dis 1:314-25.
532 2. Uttley AH, Collins CH, Naidoo J, George RC. 1988. Vancomycin-resistant
533 enterococci. Lancet 1:57-8.

534 3. Sahm DF, Kissinger J, Gilmore MS, Murray PR, Mulder R, Solliday J, Clarke
535 B. 1989. In vitro susceptibility studies of vancomycin-resistant *Enterococcus*
536 *faecalis*. Antimicrob Agents Chemother 33:1588-91. 172707
537 4. Arias CA, Murray BE. 2012. The rise of the *Enterococcus*: beyond vancomycin
538 resistance. Nat Rev Microbiol 10:266-78. 3621121
539 5. Palmer KL, Godfrey P, Griggs A, Kos VN, Zucker J, Desjardins C, Cerqueira
540 G, Gevers D, Walker S, Wortman J, Feldgarden M, Haas B, Birren B, Gilmore
541 MS. 2012. Comparative genomics of enterococci: variation in *Enterococcus*
542 *faecalis*, clade structure in *E. faecium*, and defining characteristics of *E.*
543 *gallinarum* and *E. casseliflavus*. MBio 3:e00318-11. 3374389
544 6. Lebreton F, van Schaik W, McGuire AM, Godfrey P, Griggs A, Mazumdar V,
545 Corander J, Cheng L, Saif S, Young S, Zeng Q, Wortman J, Birren B, Willems
546 RJ, Earl AM, Gilmore MS. 2013. Emergence of epidemic multidrug-resistant
547 *Enterococcus faecium* from animal and commensal strains. MBio 4. 3747589
548 7. Homan WL, Tribe D, Poznanski S, Li M, Hogg G, Spalburg E, van Embden
549 JDA, Willems RJL. 2002. Multilocus Sequence Typing Scheme for
550 *Enterococcus faecium*. Journal of Clinical Microbiology 40:1963-1971.
551 8. Lee T, Pang S, Abraham S, Coombs GW. 2019. Antimicrobial-resistant CC17
552 *Enterococcus faecium*: The past, the present and the future. J Glob Antimicrob
553 Resist 16:36-47.
554 9. Werner G, Freitas AR, Coque TM, Sollid JE, Lester C, Hammerum AM,
555 Garcia-Migura L, Jensen LB, Francia MV, Witte W, Willems RJ, Sundsfjord A.
556 2011. Host range of enterococcal *vanA* plasmids among Gram-positive
557 intestinal bacteria. J Antimicrob Chemother 66:273-82.
558 10. Freitas AR, Tedim AP, Francia MV, Jensen LB, Novais C, Peixe L, Sanchez-
559 Valenzuela A, Sundsfjord A, Hegstad K, Werner G, Sadowy E, Hammerum
560 AM, Garcia-Migura L, Willems RJ, Baquero F, Coque TM. 2016. Multilevel
561 population genetic analysis of *vanA* and *vanB Enterococcus faecium* causing
562 nosocomial outbreaks in 27 countries (1986-2012). J Antimicrob Chemother
563 71:3351-3366.
564 11. Partridge SR, Kwong SM, Firth N, Jensen SO. 2018. Mobile Genetic Elements
565 Associated with Antimicrobial Resistance. Clin Microbiol Rev 31.
566 PMC6148190

567 12. Courvalin P. 2006. Vancomycin resistance in gram-positive cocci. Clin Infect
568 Dis 42 Suppl 1:S25-34.
569 13. Wardal E, Kuch A, Gawryszewska I, Zabicka D, Hryniewicz W, Sadowy E.
570 2017. Diversity of plasmids and Tn*1546*-type transposons among VanA
571 *Enterococcus faecium* in Poland. Eur J Clin Microbiol Infect Dis 36:313-328.
572 PMC5253160
573 14. Freitas AR, Novais C, Tedim AP, Francia MV, Baquero F, Peixe L, Coque TM.
574 2013. Microevolutionary events involving narrow host plasmids influences
575 local fixation of vancomycin-resistance in *Enterococcus* populations. PLoS
576 One 8:e60589. PMC3612099
577 15. Bayliss SC, Hunt VL, Yokoyama M, Thorpe HA, Feil EJ. 2017. The use of
578 Oxford Nanopore native barcoding for complete genome assembly.
579 Gigascience 6:1-6. PMC5467021
580 16. Ping S, Mayorga-Reyes N, Price VJ, Onuoha M, Bhardwaj P, Rodrigues M,
581 Owen J, Palacios Araya D, Akins RL, Palmer KL. 2021. Characterization of
582 presumptive vancomycin-resistant enterococci recovered during infection
583 control surveillance in Dallas, Texas, USA. Access Microbiol 3:000214.
584 PMC8209702
585 17. Dutka-Malen S, Evers S, Courvalin P. 1995. Detection of glycopeptide
586 resistance genotypes and identification to the species level of clinically
587 relevant enterococci by PCR. J Clin Microbiol 33:24-7. 227872
588 18. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation.
589 Bioinformatics 30:2068-9.
590 19. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, Fookes M,
591 Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan
592 genome analysis. Bioinformatics 31:3691-3. PMC4817141
593 20. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and
594 post-analysis of large phylogenies. Bioinformatics 30:1312-3. PMC3998144
595 21. Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and
596 new developments. Nucleic Acids Res 47:W256-W259. PMC6602468
597 22. Jolley KA, Bray JE, Maiden MCJ. 2018. Open-access bacterial population
598 genomics: BIGSdb software, the PubMLST.org website and their applications.
599 Wellcome Open Res 3:124. PMC6192448

600 23. Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O, Villa L,
601 Moller Aarestrup F, Hasman H. 2014. *In silico* detection and typing of plasmids
602 using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob
603 Agents Chemother 58:3895-903. PMC4068535
604 24. Galata V, Fehlmann T, Backes C, Keller A. 2019. PLSDB: a resource of
605 complete bacterial plasmids. Nucleic Acids Res 47:D195-D202. PMC6323999
606 25. Schmartz GP, Hartung A, Hirsch P, Kern F, Fehlmann T, Muller R, Keller A.
607 2022. PLSDB: advancing a comprehensive database of bacterial plasmids.
608 Nucleic Acids Res 50:D273-D278. PMC8728149
609 26. Sullivan MJ, Petty NK, Beatson SA. 2011. Easyfig: a genome comparison
610 visualizer. Bioinformatics 27:1009-10. PMC3065679
611 27. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. 2018. HMMER
612 web server: 2018 update. Nucleic Acids Res 46:W200-W204. PMC6030962
613 28. Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions
614 in the control region of mitochondrial DNA in humans and chimpanzees. Mol
615 Biol Evol 10:512-26.
616 29. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O,
617 Aarestrup FM, Larsen MV. 2012. Identification of acquired antimicrobial
618 resistance genes. J Antimicrob Chemother 67:2640-4. PMC3468078
619 30. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the
620 reference centre for bacterial insertion sequences. Nucleic Acids Res 34:D32-
621 6. PMC1347377
622 31. McKenney PT, Ling L, Wang G, Mane S, Pamer EG. 2016. Complete Genome
623 Sequence of *Enterococcus faecium* ATCC 700221. Genome Announc 4.
624 PMC4888979
625 32. Arthur M, Molinas C, Depardieu F, Courvalin P. 1993. Characterization of
626 Tn*1546*, a Tn*3*-related transposon conferring glycopeptide resistance by
627 synthesis of depsipeptide peptidoglycan precursors in *Enterococcus faecium*
628 BM4147. J Bacteriol 175:117-27. 196104
629 33. Guerin F, Sassi M, Dejoies L, Zouari A, Schutz S, Potrel S, Auzou M, Collet A,
630 Lecointe D, Auger G, Cattoir V. 2020. Molecular and functional analysis of the
631 novel *cfr*(D) linezolid resistance gene identified in *Enterococcus faecium*. J
632 Antimicrob Chemother 75:1699-1703.

633 34. Bakthavatchalam YD, Puraswani M, Livingston A, Priya M, Venkatesan D,
634 Sharma D, Iyadurai R, Pichamuthu K, Veeraraghavan B, Mathur P. 2022.
635 Novel linear plasmids carrying *vanA* cluster drives the spread of vancomycin
636 resistance in *Enterococcus faecium* in India. J Glob Antimicrob Resist 29:168-
637 172.
638 35. van Hal SJ, Espedido BA, Coombs GW, Howden BP, Korman TM, Nimmo
639 GR, Gosbell IB, Jensen SO. 2017. Polyclonal emergence of *vanA*
640 vancomycin-resistant *Enterococcus faecium* in Australia. J Antimicrob
641 Chemother 72:998-1001.
642 36. Lee RS, Goncalves da Silva A, Baines SL, Strachan J, Ballard S, Carter GP,
643 Kwong JC, Schultz MB, Bulach DM, Seemann T, Stinear TP, Howden BP.
644 2018. The changing landscape of vancomycin-resistant *Enterococcus faecium*
645 in Australia: a population-level genomic study. J Antimicrob Chemother
646 73:3268-3278.
647 37. Valdezate S, Miranda C, Navarro A, Freitas AR, Cabrera JJ, Carrasco G,
648 Coque TM, Jimenez-Romano E, Saez-Nieto JA. 2012. Clonal outbreak of
649 ST17 multidrug-resistant *Enterococcus faecium* harbouring an Inc18-
650 like::Tn*1546* plasmid in a haemo-oncology ward of a Spanish hospital. J
651 Antimicrob Chemother 67:832-6.
652 38. Ryan L, O'Mahony E, Wrenn C, FitzGerald S, Fox U, Boyle B, Schaffer K,
653 Werner G, Klare I. 2015. Epidemiology and molecular typing of VRE
654 bloodstream isolates in an Irish tertiary care hospital. J Antimicrob Chemother
655 70:2718-24.
656 39. de Been M, Pinholt M, Top J, Bletz S, Mellmann A, van Schaik W, Brouwer E,
657 Rogers M, Kraat Y, Bonten M, Corander J, Westh H, Harmsen D, Willems RJ.
658 2015. Core Genome Multilocus Sequence Typing Scheme for High-Resolution
659 Typing of *Enterococcus faecium*. J Clin Microbiol 53:3788-97. PMC4652124
660 40. Rios R, Reyes J, Carvajal LP, Rincon S, Panesso D, Echeverri AM, Dinh A,
661 Kolokotronis SO, Narechania A, Tran TT, Munita JM, Murray BE, Planet PJ,
662 Arias CA, Diaz L. 2020. Genomic Epidemiology of Vancomycin-Resistant
663 *Enterococcus faecium* (VREfm) in Latin America: Revisiting The Global VRE
664 Population Structure. Sci Rep 10:5636. PMC7101424
665 41. Camargo IL, Gilmore MS, Darini AL. 2006. Multilocus sequence typing and
666 analysis of putative virulence factors in vancomycin-resistant and vancomycin-

667 sensitive *Enterococcus faecium* isolates from Brazil. Clin Microbiol Infect
668 12:1123-30.
669 42. McCracken M, Wong A, Mitchell R, Gravel D, Conly J, Embil J, Johnston L,
670 Matlow A, Ormiston D, Simor AE, Smith S, Du T, Hizon R, Mulvey MR,
671 Canadian Nosocomial Infection Surveillance P. 2013. Molecular epidemiology
672 of vancomycin-resistant enterococcal bacteraemia: results from the Canadian
673 Nosocomial Infection Surveillance Program, 1999-2009. J Antimicrob
674 Chemother 68:1505-9.
675 43. Moreno F, Grota P, Crisp C, Magnon K, Melcher GP, Jorgensen JH, Patterson
676 JE. 1995. Clinical and molecular epidemiology of vancomycin-resistant
677 *Enterococcus faecium* during its emergence in a city in southern Texas. Clin
678 Infect Dis 21:1234-7.
679 44. Chilambi GS, Nordstrom HR, Evans DR, Ferrolino JA, Hayden RT, Maron GM,
680 Vo AN, Gilmore MS, Wolf J, Rosch JW, Van Tyne D. 2020. Evolution of
681 vancomycin-resistant *Enterococcus faecium* during colonization and infection
682 in immunocompromised pediatric patients. Proc Natl Acad Sci U S A
683 117:11703-11714. PMC7261057
684 45. Arredondo-Alonso S, Top J, McNally A, Puranen S, Pesonen M, Pensar J,
685 Marttinen P, Braat JC, Rogers MRC, van Schaik W, Kaski S, Willems RJL,
686 Corander J, Schurch AC. 2020. Plasmids Shaped the Recent Emergence of
687 the Major Nosocomial Pathogen *Enterococcus faecium*. mBio 11.
688 `
689
690
691
692
693
694
695
696
697

698 **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

699

700 [1]Specifies the total number of isolates among the 47 Dallas isolates harboring the
701 particular pMIX group plasmids.
702 [2]The % of isolates harboring the specific plasmid group.
703 [3,4]The "% identity (DNA)" and the "% identity (aa)" specifies the % identical sites
704 calculated from the MUSCLE sequence alignment of DNA and amino acid of the *rep*
705 genes belonging to the same pMIX group, respectively.
706 [5,6] The *rep* gene encoded protein family (pfam) and the corresponding pfam accession
707 numbers, respectively.
708
709
710
711
712
713
714
715
716

717 **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	

718

719 [1]Multiple locus sequence type (MLST) of the isolates

720 [2,3]The presence of *rep* and toxin-antitoxin (TA) genes are indicated by blue boxes. The
721 pRUM-like plasmid from isolate 113 harbored *rep18a* (reference gene accession
722 AB158402) and is symbolized as “a” in the corresponding box. The *rep18* for rest of the
723 resistance plasmids were *rep18b* (reference gene accession CP018068).
724 [4]The average size of the plasmids for each plasmid group type is shown along with the
725 standard deviation (Avg. ± stdev).
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745

746 **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								

747

748 [1] Nucleotide sequences of the Tn*1546* region for the Tn*1546* groups, including more

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

750 [2]The name of the IS elements present in the Tn*1546* group are listed.

751 [3]Point mutations in the van gene cluster relative to the reference Tn*1546* (accession

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

753

754

755

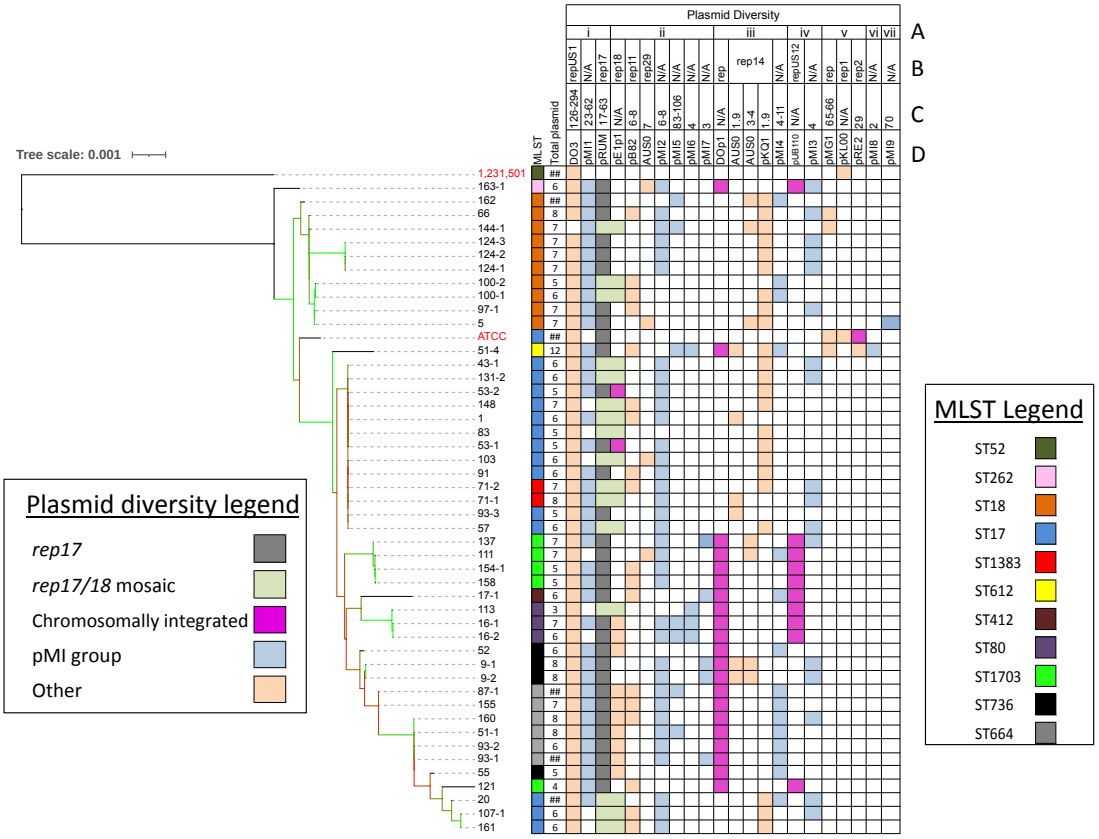


Figure 1. Phylogeny and plasmid content of Dallas *E. faecium* isolates. The phylogenetic tree of 49 *E. faecium* strains is based on an alignment of SNPs in 1706 core genes. The branches colored green, golden and red are supported by a bootstrap value of >70, 40-60, and <30 respectively. The reference 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. Plasmid groups defined in our study (pMI1 to pMI9; see main text) are highlighted in light blue color.

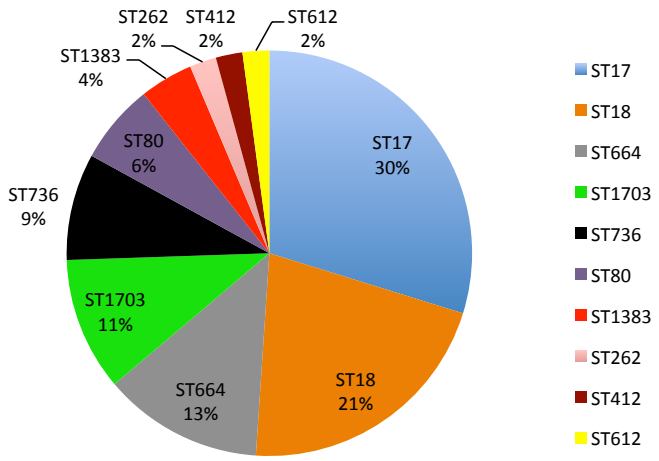


Figure 2: Sequence type (ST) distribution of Dallas *E. faecium* isolates. A total of 10 STs were observed across 47 isolates. The frequency of STs (percent of total isolates) is shown.

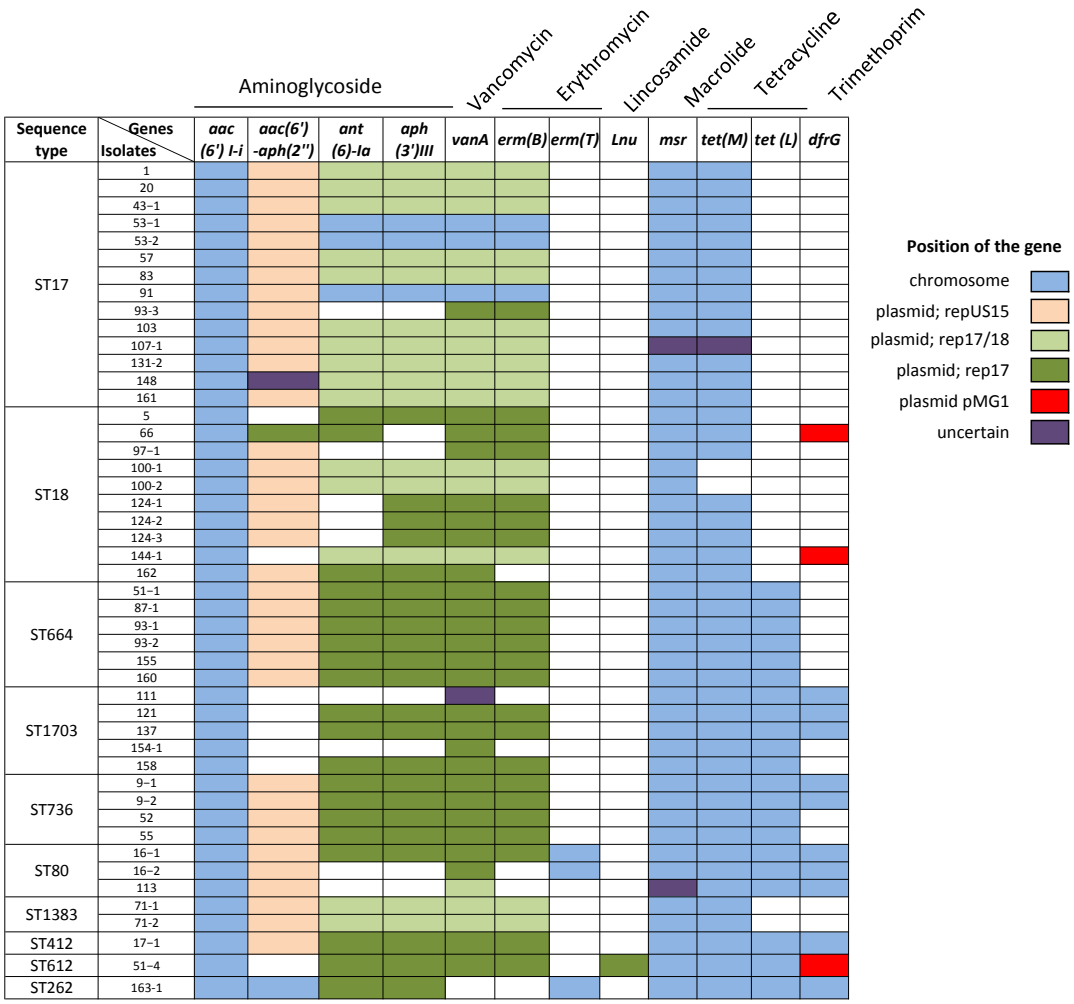


Figure 3. Distribution of predicted antibiotic resistance genes among Dallas isolate genomes. Shown are 12 resistance genes (6'-N-aminoglycoside acetyltransferase, (*aac(6')I-i*), bifunctional enzyme 6'-aminoglycoside acetyltransferase-2"-aminoglycoside phosphotransferase (*aac(6')-aph(2")*), aminoglycoside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

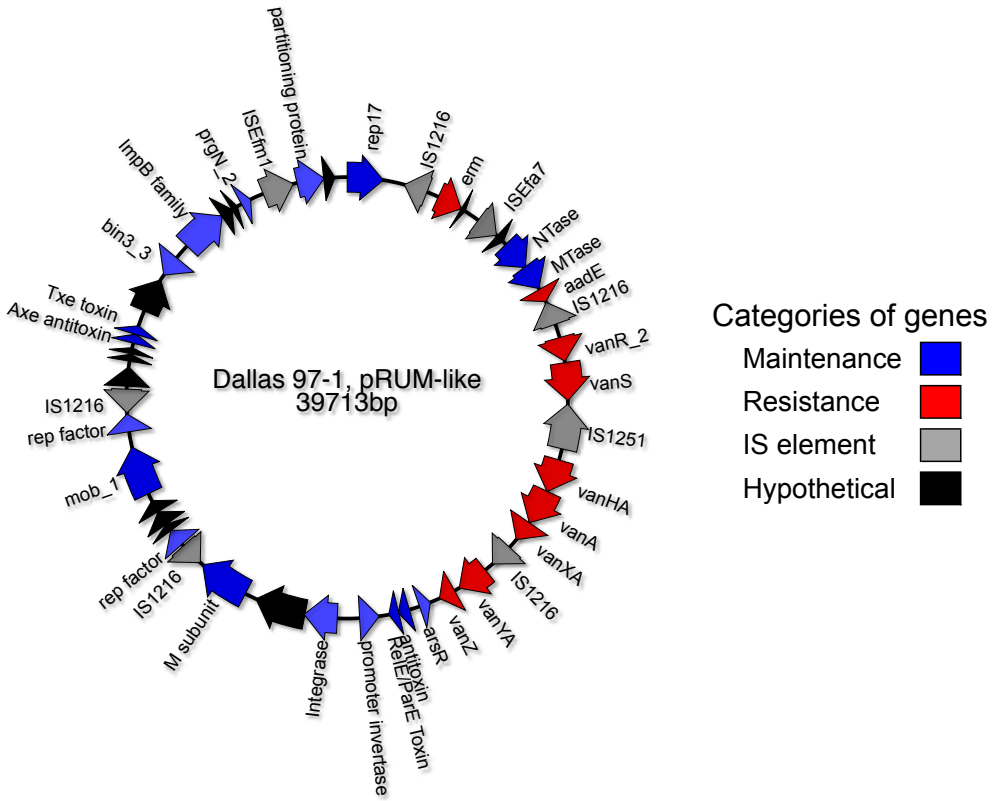
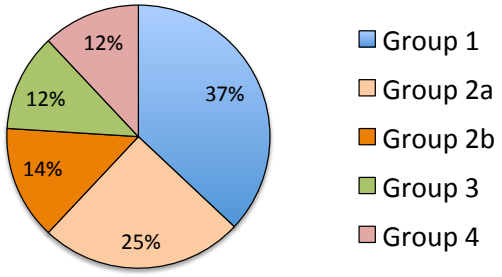


Figure 4. Genetic map of a representative Group 2a pRUM-like plasmid. The 39 kb plasmid p97_1_39kb (accession number CP066571) from the Dallas VREfm isolate 97-1 is shown. Arrows represent genes. The genes are color-coded based on their predicted activity. Maintenance genes are predicted to be responsible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

A.

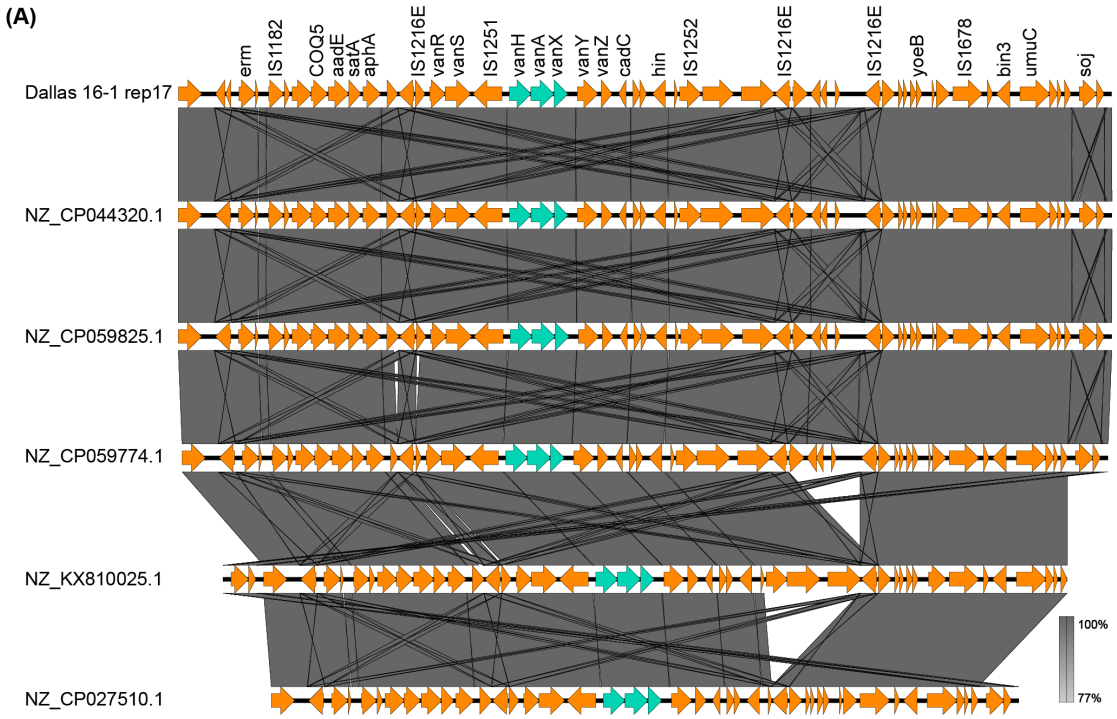
		Group 1	Group 2a	Group 2b	Group 3	Group 4
rep	*rep18*	+	-	-	-	-
	rep17	+	+	+	+	+
TA	*axe*	+	+	+	-	-
	relE	+	+	+	+	-

B.



832
833
834 **Figure 5. Classification and distribution of pRUM-like plasmid groups. A**) pRUM-
835 like plasmids were classified based on the presence or absence (indicated by a + or -,
836 respectively) of 2 replication modules (*rep17* and *rep18*) and 2 stability modules (toxin-
837 antitoxin system; $TA_{axe\text{-}txe}$ and TA_{relE}). **B)** Frequency of the pRUM-like plasmid groups
838 among 43 Dallas isolates.
839
840
841
842
843
844

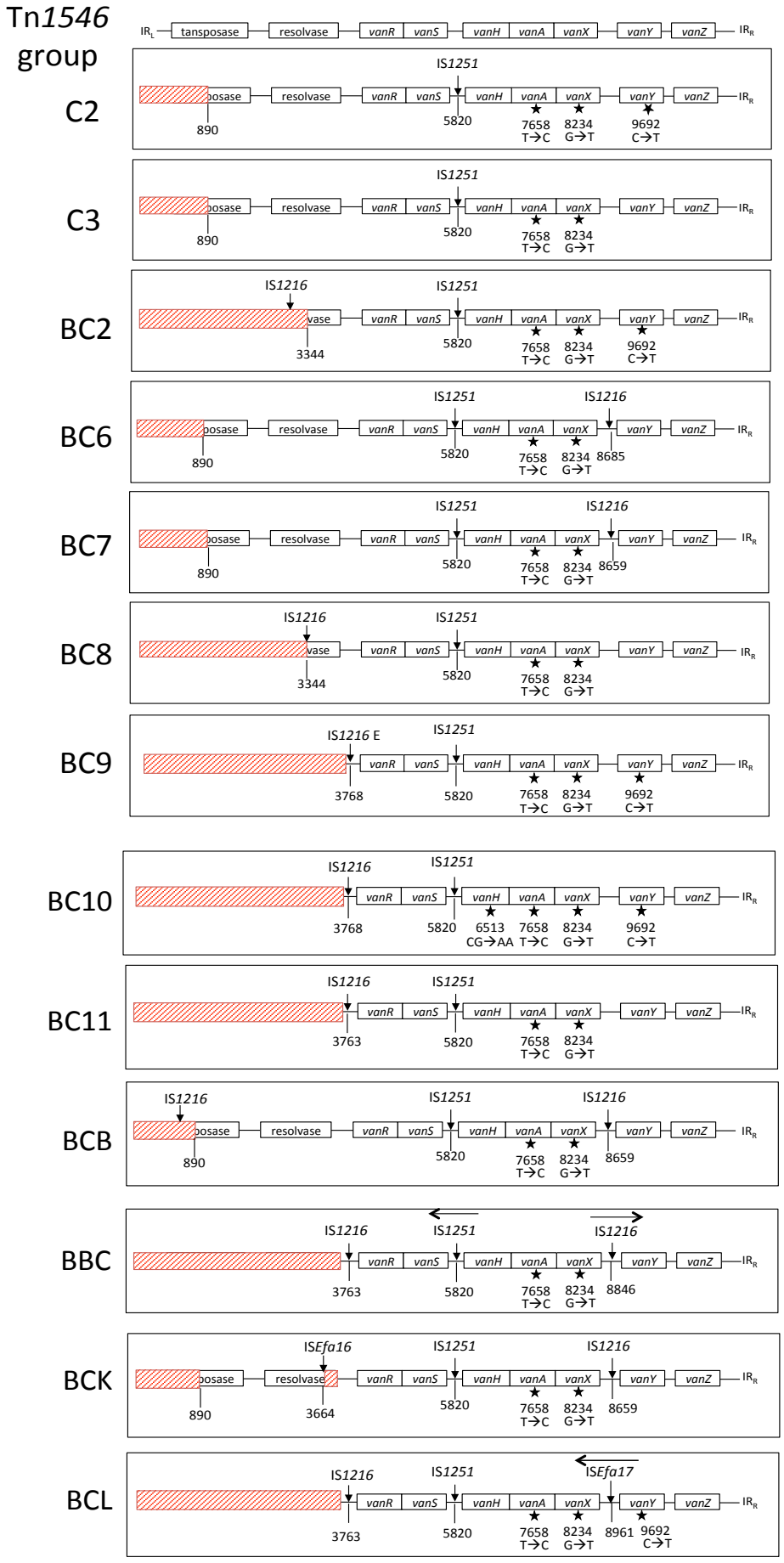
845



(B)

RefSeq Accession	Strain	ST	Size (bp)	Country of Origin	Year	Source
NZ_CP066385.1	Dallas 16-1	80	41,600	United States	2015	Rectal Swab
NZ_CP044320.1	15-307-1	17	41,619	France	2015	Rectal Swab
NZ_CP059825.1	A15023	1643	41,614	India	2019	Blood
NZ_CP059774.1	A3895	80	41,278	India	2019	Blood
NZ_KX810025.1	Efm008	80	37,639	Australia	2011	Human
NZ_CP027510.1	AUSMDU00004055	78	33,328	Australia	2015	Not Blood

846
847
848
849 **Figure 6. A cluster of pRUM-like plasmids identified from PLSDB with 99.9%**
850 **ANI.** (A) BLASTn alignment of plasmid sequences with Easyfig. (B) Data for *E.*
851 *faecium* strains and plasmids, arranged in the same order as the alignment.


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

852

853

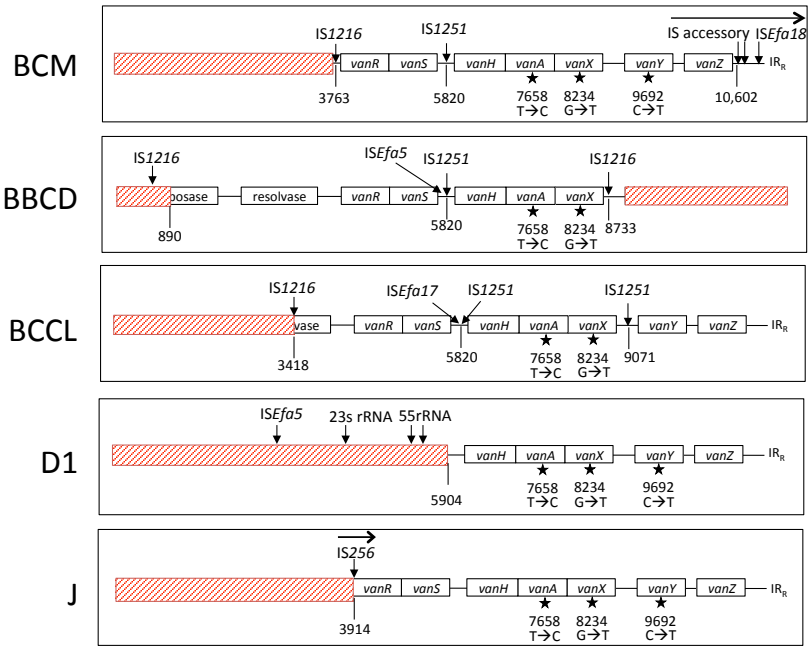


Figure 7: Tn*1546* structural variations in the Dallas VREfm isolates. A reference Tn*1546* structure that includes transposase, resolvase and *van* gene cluster flanked by IR_L and IR_R (inverted repeat left and right, respectively) is delineated at the top. The Tn*1546* groups are represented below 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.