Consideration of within-patient diversity highlights transmission pathways and antimicrobial resistance gene variability in vancomycin resistant *Enterococcus faecium*

Martin P McHugh,[1,2]* Kerry A Pettigrew,[1] Surabhi Taori,[2] Thomas J Evans,[3] Alistair Leanord,[3,4] Stephen H Gillespie,[1] Kate E Templeton,[2] Matthew TG Holden[1]

[1]School of Medicine, University of St Andrews, St Andrews, UK

[2]Medical Microbiology, Department of Laboratory Medicine, Royal Infirmary of Edinburgh, Edinburgh, UK

[3]School of Infection and Immunity, University of Glasgow, Glasgow, UK

[4]Scottish Microbiology Reference Laboratories, Glasgow, UK

*Corresponding author: Martin McHugh mpm20@st-andrews.ac.uk

ABSTRACT

Whole genome sequencing is increasingly applied to healthcare-associated vancomycin resistant *Enterococcus faecium* (VREfm) outbreak investigations using single bacterial colonies from individual patients. However, within-patient diversity could complicate transmission resolution. We sequenced 14 colonies from VREfm positive rectal swabs over a 1-month period on a haematology unit. In total, 224 isolates from 11 patients were sequenced. Carriage of 2-3 sequence types was detected in 27% of patients. Presence of antimicrobial resistance genes and plasmids was variable within isolates from the same patient and sequence type. We identified two dominant sequence types (ST80 and ST1424), with two putative transmission clusters of two patients within ST80, and a single cluster of six patients within ST1424. We

found transmission resolution was impaired using fewer than 14 colonies, and therefore there is a need to capture and consider VREfm within-patient diversity to ensure accurate resolution of transmission networks.

INTRODUCTION

Enterococcus faecium is a leading nosocomial pathogen causing opportunistic infections mostly in immunocompromised hosts. Antimicrobial resistance is a key concern, particularly against front-line anti Gram-positive agents amoxicillin and vancomycin (1). Vancomycin resistant *E. faecium* (VREfm) infections cause significant issues due to increased length of stay, cost an estimated USD200 (GBP ~167) per case per day, and confer mortality of 23-47% (2–7). In 2020, vancomycin resistance of 45.6% was reported among all *E. faecium* bloodstream isolates in Scotland, among the highest rates in Europe (8).

VREfm does not cause symptomatic gastrointestinal disease but asymptomatic intestinal carriage is common. In healthcare settings, this can increase the risk of shedding into the environment and transfer to other patients or staff, hence challenging efforts to limit the incidence of nosocomial infections in high-risk settings (9). Whole genome sequencing is increasingly applied to investigate transmission networks and identify control measures (10,11). Most whole genome sequencing (WGS) based analyses of bacterial outbreaks, however, rely on analysing single colony picks from clinical samples making the assumption that this represents the entire infecting or colonizing population within individual patients (12). However, it is increasingly recognised that within-patient diversity of bacterial populations can be significant from a clinical and infection prevention control point of view, and can influence transmission network resolution (13–19). Several studies have identified that individual patients

can carry multiple strains of *E. faecium* concurrently, but few have applied this to transmission resolution (20–24).

In this study we aimed to identify within-patient diversity of VREfm from rectal screening swabs and determine how this impacts transmission inference in a 1-month snapshot on a haematology unit. We designed a sampling strategy to reliably detect within-patient diversity, and supplemented short read and long read sequencing to generate high-quality reference genomes to identify genomic variants in the isolate collection.

RESULTS

Epidemiological Context

This study was performed over a single month in 2017 on an 18-bed unit for patients undergoing management of haematological malignancies. The unit is split into two wards (A and B). Due to the high risk of invasive infection in this patient group, VREfm rectal screening was performed on all new febrile admissions and any inpatients with febrile episodes to inform patient placement and antimicrobial administration. Given the short study timeframe there was significant overlap between patient stays with some patients moving between the two study wards or to other wards in the hospital (Figure 1). There was a policy in place for patients to be cohorted or placed in side rooms based on known colonisation or infection with high risk pathogens (including *Candida auris*, *Clostridioides difficile*, methicillin resistant *Staphylococcus aureus*, multiresistant Gram negative organisms, norovirus, respiratory viruses, varicella zoster virus, and VREfm). However, as not all rooms had ensuite facilities, risk of VREfm transmission remained. At the time of the study, pathogen surveillance systems in the hospital had not detected any suspected VREfm outbreak within the study population.

76 *Results of VREfm Screening*

77 In total, 45 rectal swabs from 27 patients were screened for VREfm. Of these, 18 samples from
78 13 patients were VREfm positive (Table 1). Additionally, three (23.1%) colonised patients
79 developed VREfm bacteraemia 9, 24, or 46 days after being identified as VREfm carriers. We
80 applied our sampling strategy to 16 rectal screens generating 224 isolates from 11 patients, and
81 five blood cultures (five isolates) from two patients yielding a total of 229 isolates. Two rectal
82 swabs and one blood culture were not available for further study. Most patients were female,
83 and the median age was 66 years, a range of primary diagnoses were represented in the
84 patient cohort (Table 1). Most colonised patients had received antibiotics in the preceding six
85 months, although only 30% had received vancomycin (Table 1).

86

87 *Simultaneous carriage of multiple VREfm strains*

88 *In silico* MLST typing using short reads from all 229 isolates showed four sequence types (ST)
89 from the hospital-associated clade A1 (25) were present in the sample cohort, dominated by
90 ST80 (n=130) and ST1424 (n=97) (Table 2). ST789 and ST1659 were identified in single
91 isolates only. Multiple STs were detected in three (27%) samples. Sample VRED06 from patient
92 P49 carried 10 (71.4%) ST80, three (21.4%) ST1424, and one (7.1%) ST789 isolate; sample
93 VRED07 from P14 carried 10 (71.4%) ST1424 and four (28.6%) ST80 isolates; sample VRED11
94 from P50 carried 13 (92.9%) ST1424 and one (7.1%) ST1659 isolate. A further rectal swab
95 sample from P49 collected two days after VRED06 contained only ST1424, and a blood culture
96 collected nine days later also contained ST1424. P9 had three rectal swab samples collected
97 over an 11 day period and had positive blood cultures one month later, all samples contained
98 ST80 only.

99

100 *Generation of high quality reference genomes*

101 As 99% of isolates grouped into only two STs we next sought to identify relationships within
102 these two sequence types. To generate strain-specific reference genomes for further analysis
103 one ST80 (VRED06-10) isolate and one ST1424 (VRED06-02) isolate were selected for long
104 read sequencing. Both isolates were from P49 and were from the same swab sample collected
105 on study day 15. Complete genomes were assembled and polished using the Trycycler pipeline
106 generating a single circular chromosome and 5-7 plasmids for each isolate (Supplementary
107 Table 1). BUSCO showed 99.5% complete core genes and indel analysis showed <4% coding
108 sequences were significantly shorter than the reference match, within the expected range of
109 truncated genes due to evolutionary processes (26). These results show the assemblies were
110 complete and contained very few errors.
111
112 *Genomic population structure of VREfm isolates suggests recent transmission events*
113 The chromosomes of the two strain-specific genome assemblies were used as references for
114 short read mapping within each sequence type. Within-patient diversity was low when isolates
115 of the same MLST ST were compared, generally differing by zero SNPs and a maximum
116 pairwise difference of 4 SNPs (Table 2). Where within-patient isolates were from different STs,
117 SNP differences were significantly higher as expected (490-4,400 SNPs, Table 2). Similarly,
118 detection of insertions, deletions, and plasmids were usually shared in isolates from the same
119 patient. However, the presence of DEL3 (12 bp non-coding deletion) and DEL4 (11 bp deletion
120 in a solute binding protein) were variable within 24 ST80 isolates from P20 with 0-2
121 differentiating SNPs (Figure 2). In isolates from P9 p1_VRED06-10 and p3_VRED06-10 were
122 variably detected despite most genomes having no differentiating SNPs (Figure 2). Where
123 multiple samples from the same patient were collected over time we found low (0-3 SNPs)
124 accumulation of SNPs and no pattern in the prevalence of other genomic variants, reflecting the
125 short study time period.

126 The ST80 genomes formed a well-structured population with five clear clusters each
127 separated by >10 SNPs (Figure 2). Within the clusters isolates were closely related, differing by
128 0-2 SNPs, and were mostly derived from individual patients. However, two clusters included
129 isolates from two different patients (patients P7 and P33, and P2 and P9), suggesting two
130 recent transmission events. All the reference plasmids were detected in the P7 and P33 isolates
131 and there was a difference of two insertions in three isolates. As mentioned above there was
132 significant variation in detection of p1_VRED06-10 and p3_VRED06-10 plasmids within P9
133 isolates, although in P2 isolates all plasmids were detected.
134 Mapping of the ST1424 isolates showed a much more homogeneous population than in
135 ST80 (Figure 3). Of the 97 ST1424 isolates, 69 had no SNPs and the remaining 28 had 1-2
136 SNPs differentiating them from the rest of the collection. The SNPs that were detected did not
137 lead to any clear clustering of isolates, except for the 14 isolates from P6 which all carried a
138 SNP in a penicillin binding protein which differentiated them from the other ST1424 isolates.
139 Two of the P6 isolates had further independent SNPs (one each) and another one isolate had
140 lost p1_VRED06-02. Two isolates from P49 collected on day 17 shared a SNP in a penicillin
141 binding protein as well as a 13 bp deletion in a non-coding region. The same variants were
142 detected in a blood culture isolate collected 13 days later. Of note, the specific SNP in these
143 three P49 isolates was in the same penicillin binding protein gene as the P6 isolates, but a
144 different region. No insertions were detected in the ST1424 collection, and of the six deletions
145 found five were only in isolates from P49. Apart from the loss of p6_VRED06-02 mentioned
146 above, the only other plasmid not detected was p1_VRED06-02 in five isolates from three
147 patients.
148
149 *Analysis of multiple VREfm colonies supports transmission resolution*

150 A transmission network was then constructed taking account of the phylogenetic placement of
151 all 14 colony picks in each sample (Figure 4). The network supports transmission of ST80
152 between P2 and P9, and between P7 and P33, although P20 appears to be a singleton not
153 linked to transmission within this patient cohort. Considering patient movements, P9 was
154 admitted to Ward B prior to the start of the study period and screened positive for VREfm on day
155 13 of their admission, P2 screened positive on the same day although this was their first day of
156 admission to Ward B (Figure 1). Both P2 and P9 were admitted through Ward A prior to being
157 placed on Ward B, although no shared rooms or bed spaces were identified (Supplementary
158 Figure 1). The directionality or location of this putative transmission event cannot be
159 ascertained. P7 screened positive on admission to Ward B. P33 was admitted from an external
160 Health Board to Ward B, screened negative on day 2 of their admission, then screened positive
161 on day 10 of admission (6 days after P7). No shared rooms or bed spaces were identified for P7
162 and P33, although given the timings of positive screens it is likely that ST80 VREfm was
163 transmitted from P7 to P33 on Ward B (Figure 1, Supplementary Figure 1). P20 screened
164 negative towards the end of a 32 day stay on Ward B, was discharged for 13 days and then
165 readmitted to Ward B and screened positive on admission, suggesting they may have become
166 colonised outside of the hospital.
167 Patients carrying ST1424 all clustered together with P34 having strong links to all other
168 patients, including likely direct transmission to P6 (Figure 4). P34 was admitted to Ward B from
169 an external Health Board and screened positive with ST1424 on day 6 of admission, 6 days
170 later P49 screened positive. P49 was on Ward B two days prior to P34 but screened negative 3
171 days before P34 was found to be colonised, suggesting P34 was the source of ST1424
172 transmission to P49. Soon after, P14 was found to be colonised with ST1424 and ST80 on day
173 2 of admission to Ward B but transmission from P49 is unlikely as ST80 isolates in both patients
174 were unrelated and P49 had an additional ST789 strain. It is possible that P14 and P49 had

175 superinfection of ST1424 on top of existing VREfm colonisation. Nine days later, P24 screened
176 positive with ST1424 on Ward B on their seventh day of admission. P24 shared time on the
177 ward with all three previously identified carriers, based on genetic similarity of all sequenced
178 isolates P34 is the most parsimonious source of transmission (Figure 3). P6 and P50 were both
179 admitted to Ward A on the same day late in the study period, P50 screened positive with
180 ST1424 and ST1659 on day 2 and P6 screened positive with ST1424 only on day 6. From the
181 epidemiology it may be assumed that P6 was infected by P50 however, sequencing suggests
182 the two ST1424 populations may be derived from different hosts, with P6 isolates having all
183 acquired a single SNP and P50 isolates having multiple different SNPs and all lacking the
184 p6_VRED06-02 plasmid (Figures 3 and 4). P6 and P50 were the only patients on Ward A to
185 acquire ST1424, the route of transmission between the two wards is unclear. None of the
186 patients with ST1424 shared a room or used the same bed space previously used by an
187 identified ST1424 positive carrier during their stay (Supplementary Figure 1).
188 Performing the analysis again with three, five, or ten genomes per screening sample did
189 not change the transmission inference for ST80 isolates (although higher confidence was
190 achieved with five or more genomes), but there was significant variability in the ST1424 cluster
191 due to low diversity in the genomes (Supplementary Figure 2). Within ST1424 it was only
192 possible to reliably link all patients and give high confidence that P34 was the main driver in the
193 cluster with 14 isolates.
194
195 *Plasmids were mostly ST-specific*
196 Focusing on the plasmids identified in the reference isolates, VRED06-02 (ST1424) contained
197 seven plasmids, and VRED06-10 (ST80) contained five plasmids. Plasmids in the two isolates
198 were not particularly similar, suggesting there was limited sharing of plasmids between the two
199 isolates within P49 at the time of sampling (Supplementary Table 2).

200 We then sought to identify carriage of similar plasmids in the entire isolate collection by
201 mapping the short reads to each assembled plasmid (Supplementary Table 3). This showed
202 that for the most part plasmids were ST-specific as there were few examples of ST1424 isolates
203 carrying plasmids from the ST80 reference, and *vice versa*. However, all ST80 isolates from P7
204 and P33 carried p7_VRED06-02 from ST1424, and almost all isolates appeared to carry
205 p4_VRED06-10. p4_VRED06-02 and p4_VRED06-10 had the closest pairwise identity
206 (Supplementary Table 2), with mash distance of 0.05 and average nucleotide identity of 89%
207 over 4.3 kb. As p4_VRED06-10 is slightly smaller, we believe the hits against the ST1424
208 isolates here are due to cross-mapping of reads from the related p4_VRED06-02. P7_VRED06-
209 02 carriage in P7 and P33 does not look like a false positive as it is unrelated to others in the
210 collection (Supplementary Table 2), but no close links to any ST1424 positive patients were
211 identified for P7 and P33 (Figure 4), so the source of this plasmid is unclear. We also found the
212 single ST1659 isolate from P50 had an identical plasmid profile to 12 ST1424 isolates in the
213 same sample, and the single ST789 isolate from P49 had the same plasmid profile as 10 ST80
214 isolates in the same patient sample.
215
216 *AMR gene load differs between closely related isolates*
217 We next sought to determine the variability of AMR genes within the collection (Figure 5). In
218 total 13 AMR genes were detected with three (23.1%) present in all isolates, five (38.5%) genes
219 (*ant(9)-Ia*, *dfrG*, *erm(A)*, and *tet(M)*) only in ST1424 isolates and two (15.4%) genes (*ant(6)-Ia*
220 and *tet(S)*) found only in ST80 isolates. Tetracycline resistance gene *tet(M)* was found in 62.2%
221 of ST1424 and ST1659 genomes and the aminoglycoside resistance gene *aac(6')-aph(2'')* was
222 found in 69.9% of all genomes (Table 3 and Figure 5).

223 *tet(M)* was identified on the chromosome of VRED06-02, as part of Tn*6944* first
224 identified in *C. difficile* (Supplementary Figure 3A). Excision of Tn*6944* is likely responsible for
225 variable presence of *tet(M).*
226 *aac(6')-aph(2'')* was present on p1_VRED06-02 (ST80) and p1_VRED06-10 (ST1424).
227 In p1_VRED06-02, two copies of *aac(6')-aph(2'')* were surrounded by insertion sequences
228 IS*256*, IS*1216*, and IS*3,* potentially providing multiple mechanisms of excision. *aac(6')-aph(2'')*
229 was not detected in any ST80 isolates that were p1_VRED06-02 negative, although only 60.2%
230 (n=65) of isolates that carried this plasmid also carried *aac(6')-aph(2'')*, highlighting that
231 although the cassette was linked to p1_VRED06-02 it may have been excised in some cases. In
232 ST1424 p1_VRED06-10, *aac(6')-aph(2'')* was surrounded by two copies of IS*256* and carried
233 genes including *erm(B)* with similarity to Tn*6218* first identified in *C. difficile*, although the
234 transposition machinery was missing (Supplementary Figure 3B) (27). In ST1424 *aac(6')-*
235 *aph(2'')* was detected in 97.8% (n=90) isolates with p1_VRED06-10 but was also found in four
236 isolates that lacked this plasmid (Table 3 and Figure 5). Short read assemblies were too
237 fragmented to fully resolve the location of the gene in these four cases, but three of the
238 assemblies did show *aac(6')-aph(2'')* in association with an IS*3* gene suggesting it had
239 mobilised to another genetic element within the cell.
240 The tetracycline resistance gene *tet(L)* was identified in a single ST1659 isolate, the
241 gene was co-located with *tet(M)* on a 30 kb contig that was similar to Tn*6248* from *E. faecium*
242 over ~19 kb (Supplementary Figure 3C).
243
244 **DISCUSSION**
245 By taking account of within-patient diversity in VREfm carriage populations we identified
246 unknown links between patients that could supplement efforts to control transmission within
247 hospitals. We also show that diversity exists not just at the level of SNPs – AMR gene

248 presence/absence, indels, and plasmid presence all vary within and between patients. The
249 relative importance of each of these layers of diversity for understanding bacterial transmission
250 and evolution are key for future WGS based analyses.
251
252 *Identified transmission clusters*
253 We identified an ST1424 transmission cluster involving six patients and two ST80
254 clusters involving two patients (Figure 4). The ST80 population appeared relatively stable
255 otherwise, with three clusters of isolates belonging to individual patients. However, in P14 and
256 P49 where ST1424 and ST80 isolates were both present in carriage samples we did not detect
257 transmission of ST80 suggesting these individuals were infected by an ST1424 on top of
258 existing VREfm carriage, or a wide transmission bottleneck and acquisition of both lineages at
259 the same time from an undetermined source.
260 Power calculation determined 14 colonies to be the minimum needed to sequence per
261 sample to ensure reliable detection of a variant making up 20% of the within-patient population.
262 This successfully identified minor population variants in 27% of patients. We also performed
263 transmission inference with three, five, and ten isolates per rectal sample and found that the
264 strength and accuracy of transmission inference increased as the number of isolates per patient
265 increased (Figure 4 and Supplementary Figure 2). Sequencing more than 14 colonies would
266 further improve the detection of minor variants but would also increase the costs of sequencing
267 and turnaround time. Our findings show that sequencing single colonies without considering
268 within patient diversity in suspected VREfm outbreaks will miss some putative transmission
269 events, but the implementation of sequencing multiple colonies into routine use will require
270 careful thought and validation as there was low within-patient diversity in most cases in our
271 collection. Alternatively, strain-resolved metagenomics directly on clinical samples or
272 sequencing sweeps of selective culture growth may be more feasible in future (28–30). Further

273 work is required to determine the optimum sampling strategy to support infection prevention and
274 control investigations in healthcare settings.
275 We identified three (27%) patients that carried two or three different strains at the same
276 time (Table 2). This in line with recent studies that identified up to half of patients carry 2-4
277 different *E. faecium* strains and within-patient diversity varies over time, with identical AMR
278 elements passing between unrelated strains within patients (20,24,31,32). These findings
279 highlight the challenges of tracking and controlling highly variable *E. faecium* populations within
280 healthcare settings, particularly when relying on single colony picks and only short read
281 sequencing which cannot fully resolve most plasmids.
282 We found very little SNP-based diversification over time in our study, with ST1424
283 isolates often being identical to isolates collected weeks earlier, although we did identify a
284 defining SNP in the P6 isolates collected late in the study (Figure 3). Estimates of diversification
285 rates in *E. faecium* from national single colony sampling suggest 7 mutations per year (33),
286 although other studies of longitudinal within-patient diversification have estimated higher rates of
287 12.6 – 128 mutations per year (20,22,23). The low SNP diversity identified in our one-month
288 collection of carriage isolates is in keeping with these mutation rates, although it is perhaps
289 unexpected that in the four P9 bloodstream isolates only one SNP was detected in one isolate
290 compared to rectal samples collected 35-37 days previously (Figure 2). The low SNP level
291 diversity leads to difficulty in reliably inferring specific transmission events within ST1424, as
292 multiple potential transmission events are plausible, as reflected in the transmission network
293 (Figure 4).
294
295 *Variability in AMR and Mobile Elements*
296 We identified some variability of AMR gene carriage within patients, probably related to
297 loss of mobile elements (Table 3 and Figure 5). However, the impact on phenotype is unclear –

298 all isolates carried *aac(6')-Ii* and *aph(3')-III* which together confer high level resistance to the
299 clinically relevant aminoglycosides amikacin and gentamicin, so the loss of *aac(6')-aph(2'')* may
300 be more efficient for the cell without an overt phenotypic change in minimum inhibitory
301 concentration. *tet(M)* was present in 62% of ST1424 genomes but no other tetracycline
302 resistance genes were found in these isolates, so this variability could have impacts on clinical
303 reporting depending on which colony is selected for antibiotic susceptibility testing. However,
304 tetracyclines are not generally used for direct treatment of enterococcal human infections so the
305 clinical impact of this variability will be limited. Similar variable presence of the vancomycin
306 resistance element within patients has been described elsewhere and could lead to inadvertent
307 use of vancomycin even when the patient harbours a resistant population (20,21,34,35).
308 We identified a linear plasmid (p2_VRED06-10) carrying *vanA* in our ST80 reference
309 genome. Boumasmoud *et al* (36) recently described the linear plasmid pELF_USZ in VREfm
310 from Switzerland that carried an operon conferring the ability to utilise the human gut mucin N-
311 acetyl-galactosamine; however, p2_VRED06-02 did not carry this operon. The advent of long
312 read sequencing has allowed the increasing identification of linear plasmids, and their impact on
313 niche adaptation remains to be fully elucidated (36–40).
314 We recognise some limitations. Around 60% of *E. faecium* carriers in hospitals can be
315 linked to nosocomial transmission either from other patients or reservoirs in the hospital
316 environment (31,41–44). Our study did not include environmental samples, although patients
317 were mostly located in individual rooms. Bathroom facilities were shared posing a significant
318 environmental reservoir for VREfm. Also, we relied on direct plating to solid VREfm screening
319 agar for inclusion in our study. Previous studies have shown a sensitivity of 58-96% for this
320 approach, rising to 97-100% with a pre-enrichment step (45–47). In our study, P49 carried
321 ST1424, ST80, and ST789 on their first positive screen having been negative for VREfm 11
322 days previously, suggesting either infection with a mixed population from another host not

323 detected by our sampling strategy or an initial false negative VREfm screen. Our focus on
324 VREfm is another potential limitation as vancomycin sensitive *E. faecium* may contribute to
325 transmission clusters causing significant disease and could gain/lose the resistance mechanism
326 during transmission as described by Raven *et al* (33) – although isolates in the identified
327 clusters were collected over years rather than weeks in our study. Despite these limitations, we
328 believe our study are applicable to current practice as surveillance systems are mainly focussed
329 on significant antimicrobial resistance phenotypes such as vancomycin resistance. Routine,
330 prospective, genomic surveillance over longer periods than our study may be more suited to a
331 resistance agnostic approach to understanding nosocomial spread of pathogenic bacteria (31).
332 Our findings suggest that sequence-based surveillance of VREfm screening samples
333 would be useful in identifying outbreaks proactively. A proactive approach would avoid large
334 outbreaks of infection, reduce costs of ward closure and the clinical impact of invasive disease
335 (48–51). In our setting an outbreak of VREfm was suspected 3 weeks after the study collection
336 period when P9 and P49 developed bloodstream infection concurrently but this was many
337 weeks after VREfm transmission had likely occurred (Figure 1 and Figure 4). Due to the
338 retrospective nature of our study, we were not able to use the findings from sequencing to
339 directly influence patient care.
340
341 **CONCLUSION**
342
343 We have shown that sequencing 14 VREfm carriage isolates per patient improved resolution of
344 possible transmission pathways. This has implications for outbreak investigations which may
345 miss possible transmission events if only single colonies are investigated. We also show the
346 utility of long read sequencing to supplement routine short read sequencing to generate high-
347 quality, closed reference genomes and allow identification of non-SNP variation such as indels,

348 AMR genes, and plasmid loss/gain events which can be used for fine-level discrimination within
349 suspected bacterial outbreaks.
350
351 **MATERIAL AND METHODS**
352
353 *Bacterial Isolates*
354 Screening for VREfm intestinal carriage was performed by taking rectal swabs as part of
355 admission sampling pre-chemotherapy/transplant, and on all inpatients on the haematology unit
356 developing febrile neutropenia (defined as neutrophils <0.9×10^9/l or <1.0×10^9/l and falling after
357 chemotherapy, plus body temperature ≥38°C). Swabs were plated to CHROMID® VRE agar
358 (bioMérieux, Marcy-l'Étoile, France) and incubated at 37°C in air. Plates were examined for
359 growth at 24 and 48 hours. Growth of purple colonies were presumed VREfm and a single
360 colony was subcultured to Columbia blood agar plates (Oxoid, Basingstoke, UK) with a 5 μg
361 vancomycin disc. Species identification was confirmed by MALDI-TOF on a Microflex instrument
362 (Bruker, Billerica, USA), and the first positive isolate per patient had antimicrobial susceptibility
363 determined on the VITEK2 system AST-P607 card (bioMérieux). Once confirmed as VREfm, all
364 purple colonies were removed from the initial screening plate within 24 hours and stored at -
365 80°C in a Microbank cryovial (Pro-Lab Diagnostics, Birkenhead, UK). Any VREfm isolated from
366 clinical samples (blood, wounds, or urine samples) within 60 days of the patient being identified
367 as a VREfm carrier were also stored. Electronic records were reviewed for demographic and
368 inpatient stay metadata. Patient movements were visualised with HAIviz v0.3
369 (https://haiviz.beatsonlab.com/). This work used excess diagnostic material and was approved
370 by the NHS Scotland BioRepository Network (Ref TR000126) and University of St Andrews
371 Research Ethics Committee (Ref MD12651).
372

373 *DNA Extraction*

374 Two cryobank beads were used to inoculate a CHROMID® VRE agar plate and incubated for
375 20-24 hours at 37°C in air. Fourteen purple colonies were selected at random, each inoculated
376 into 5 ml of brain heart infusion broth (Oxoid) and incubated at 37°C for 20-24 hours. Bacterial
377 cells were pelleted by centrifugation (3000 g, 10 min), resuspended in 400 µl Buffer P1 (Qiagen,
378 Hilden, Germany), and 200 µl used for DNA extraction.

379 Cells were lysed with 20 µl lysozyme (100 mg/ml) (Sigma-Aldrich, Dorset, England) at
380 37°C for 60 min, followed by 30 µl Proteinase K (Qiagen) at 56°C for 60 min, and RNA was
381 digested with 4 µl RNase A (Qiagen) at 37°C for 45 min. DNA was extracted from lysates on the
382 QiaSymphony instrument using the DNA Mini kit (Qiagen) and protocol
383 Tissue_HC_200_V7_DSP. DNA concentration was determined from 2 µl extract with the Qubit
384 BR Kit (ThermoFisher Scientific, Basingstoke, UK) on a Qubit 3.0 fluorometer, extract purity was
385 assessed with a Nanodrop 1000 (ThermoFisher Scientific).

386 Isolates were processed in batches of 24, with every 8th position being 5 ml of
387 uninoculated BHI as a negative control to identify possible carry-over contamination. Negative
388 controls were processed as above and assessed for purity using the Qubit HS Kit
389 (ThermoFisher Scientific) in triplicate. Negative controls were accepted if DNA concentration
390 was <0.5 ng/µl in all replicates.

391

392 *Genome Sequencing*

393 Short read libraries were prepared using the Nextera XT kit (Illumina, San Diego, USA) and
394 sequenced with a MiSeq instrument (Illumina) using 300 bp paired-end reads on a 600-cycle v3
395 reagent kit.

396 Extracts from two isolates representative of the main clades identified in the collection
397 were selected for long read sequencing. Long read libraries were generated with the LSK109

398 Ligation Sequencing Kit (Oxford Nanopore Technologies, Oxford, UK) and sequenced for 8h
399 using an R9.4 flow cell on a GridION sequencer (Oxford Nanopore Technologies). High
400 accuracy real time basecalling was performed in MinKNOW v19.12.6 (Oxford Nanopore
401 Technologies).
402 Sequence data from this study has been deposited in the NCBI under BioProject
403 accession number PRJNA877253 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA877253).
404
405 *Sequence Assembly and Mapping*
406 Long reads were filtered to remove <1000 bp fragments using Nanofilt v2.7.1 (52), then trimmed
407 with Porechop v0.3.2 (https://github.com/rrwick/Porechop) to remove adapters and split chimeric
408 reads. Genome assemblies were generated with Trycycler v0.3.3 (53). Trycycler makes multiple
409 long read assemblies from read subsets and produces a consensus long read assembly which
410 is then polished with long and short reads. Four assemblies were generated using Flye v2.8.1,
411 Redbean v2.5, Raven v1.1.10, and Miniasm v0.1.3 (54–57), giving 12 assemblies in total. The
412 final consensus assembly was polished with Medaka v0.11.5
413 (https://github.com/nanoporetech/medaka) and Pilon v1.23 (58), the latter for 2-3 cycles until no
414 further changes were made. Assembly quality was assessed with assembly-stats v1.0.1
415 (https://github.com/sanger-pathogens/assembly-stats), Ideel (https://github.com/phiweger/ideel)
416 and Busco v4.1.4 (59).
417 Hybrid assemblies were annotated with Prokka v1.14.6 (60). Abricate v1.0.1
418 (https://github.com/tseemann/abricate) was used to identify matches to the ResFinder,
419 VirulenceFinder, and PlasmidFinder databases in the assembled contigs (61–64). Putative
420 prophages were identified with PHASTER (65), and recombination blocks were identified with
421 Gubbins v2.4.1 (66).

422 Short reads were quality trimmed with Trimmomatic v0.32 (67) and settings
423 MAXINFO:200:0.4. MLSTs were determined from the short reads using SRST2 v0.2.0 (68) and
424 the *E. faecium* pubMLST database (69). For mapping to reference genomes a core alignment
425 was generated by mapping short reads with Snippy v4.6.0 (https://github.com/tseemann/snippy)
426 and masking all putative transposases, complete prophage regions, and recombination blocks.
427 For mapping to reference genomes, a core alignment was generated by mapping short reads
428 with Snippy v4.6.0 (https://github.com/tseemann/snippy) and masking all putative transposases,
429 complete prophage regions, and recombination blocks. Non-ACGT bases were converted to N
430 with snippy-clean and then the variant sites were retained with snp-sites v2.5.1 (70). Maximum-
431 likelihood phylogenies were constructed with IQTree v2.0.3 from the core SNP alignment with
432 automatic model selection and 1000 ultrafast bootstraps (71–73). Phylogenies were visualised
433 with iTOL (74). Short read assemblies were generated with Unicycler v0.4.8 (75) and searched
434 for antimicrobial resistance genes using Abricate and the ResFinder database.
435
436 *Transmission Network Inference*
437 Short reads were mapped to the VRED06-10 ST80 reference chromosome with Snippy, the V24
438 *E. faecium* ST80 genome (Accession CP036151) was included as an outgroup. A core
439 alignment was generated as above then a posterior set of phylogenies were generated with
440 MrBayes v3.2.7 (76). Two MCMC runs consisting of four coupled chains were run for 5,000,000
441 generations and sampled every 5000 generations. At the completion of the run the standard
442 deviation of split frequencies was 0.013, the log likelihood was stable, and the effective sample
443 size of the prior, likelihood, and all numerical parameters was >800. A random sample of 100
444 trees from the MrBayes posterior set were used as input for Phyloscanner v1.6.6 (77).
445 Parsimony reconstruction was performed in Phyloscanner using the Sankoff algorithm and *k*
446 parameter of 281494.5, equivalent to a within patient diversity threshold of 10 SNPs. The host

447 relationship summary was then used to construct a transmission network in Cytoscape v3.9.0
448 (78), only edges with complex or transmission state and >0.5 probability were visualised.
449
450 *Plasmid Detection*
451 As described by Pinholt *et al* (37), to detect plasmids in the entire collection, short reads were
452 mapped to plasmids from the two complete assemblies using Snippy. Plasmids were
453 considered present if ≥85% bases were called with <20 SNPs/1000 bp.
454
455 *Statistical Analysis*
456 To determine the optimal number of colonies to analyse for within-sample diversity a power
457 calculation was performed as described by Huebner er al (79):
458
459 $$q \wedge n = (1 - P)$$
460
461 Where q = 1 – concentration of organisms, ^ = exponentiation operator, n = number of colonies
462 sequenced, and P = probability of finding one or more variants.
463 Population variants were considered distinct if they differed by >10 chromosomal SNPs.
464 Moradigaravand *et al* (20) have shown within-patient populations of VREfm can harbour
465 minority variants at 20-50% of the total population based on sequencing ~5 colonies. To detect
466 a variant at 20% of the population with 95% confidence would require 14 colonies per sample so
467 this was selected for analysis of faecal population diversity in this study. However, more in
468 depth sequencing of blood culture isolates showed much lower diversity – no minor population
469 variants from ~10 colonies (20) – therefore single colonies from blood culture samples were
470 sequenced in this study to identify the sequence type causing invasive disease.

Presence/absence matrices of AMR genes were generated in R v4.0.5 using ggplot2 and patchwork packages (80–82).

ACKNOWLEDGEMENTS

This work was supported by the Chief Scientist Office (Scotland) through the Scottish Healthcare Associated Infection Prevention Institute (Reference SIRN/10). The authors acknowledge the Research/Scientific Computing teams at The James Hutton Institute and NIAB for providing computational resources and technical support for the "UK's Crop Diversity Bioinformatics HPC" (BBSRC grant BB/S019669/1), use of which has contributed to the results reported within this paper. Bioinformatics and Computational Biology analyses were further supported by the University of St Andrews Bioinformatics Unit which is funded by a Wellcome Trust ISSF award [grant 105621/Z/14/Z].

AUTHOR CONTRIBUTIONS

M.P.M., S.H.G., K.E.T., and M.T.G.H. conceived the project. S.H.G., K.E.T., and M.T.G.H. supervised the project. M.P.M. and K.A.P. performed genome sequencing. M.P.M. performed data analysis. M.P.M. and M.T.G.H. wrote the manuscript. M.T.G.H., S.H.G., T.J.E., and A.L. secured funding for the project. All authors reviewed and approved the manuscript.

COMPETING INTERESTS

The authors declare no competing interests

496 **ETHICS DECLARATIONS**

497

498 Access to isolates and clinical data was approved by the NHS Scotland Biorepository Network

499 (Ref TR000126) and University of St Andrews Research Ethics Committee (Ref MD12651).

500

501 **REFERENCES**

502

503 1. Arias CA, Murray BE. The rise of the Enterococcus: beyond vancomycin resistance. Nat
504 Rev Microbiol. 2012;10:266–78.

505 2. WHO. WHO publishes list of bacteria for which new antibiotics are urgently needed
506 [Internet]. 2017 [cited 2022 Apr 26]. Available from: https://www.who.int/news/item/27-02-
507 2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed

508 3. CDC. Antibiotic Resistance Threats in the United States, 2019. Atlanta, GA: U.S.
509 Department of Health and Human Services, CDC; 2019.

510 4. Ong DSY, Bonten MJM, Safdari K, Spitoni C, Frencken JF, Witteveen E, et al.
511 Epidemiology, management, and risk-adjusted mortality of ICU-acquired enterococcal
512 bacteremia. Clin Infect Dis. 2015;61(9):1413–20.

513 5. Billington EO, Phang SH, Gregson DB, Pitout JDD, Ross T, Church DL, et al. Incidence,
514 Risk Factors, and Outcomes for Enterococcus spp . Blood Stream Infections: A Population-
515 Based Study. Int J Infect Dis. 2014;26:76–82.

516 6. Dik JWH, Dinkelacker AG, Vemer P, Lo-Ten-Foe JR, Lokate M, Sinha B, et al. Cost-
517 analysis of seven nosocomial outbreaks in an academic hospital. PLoS ONE. 2016;11(2):1–
518 7.

519 7. Lloyd-Smith P, Younger J, Lloyd-Smith E, Green H, Leung V, Romney MG. Economic
520 analysis of vancomycin-resistant enterococci at a Canadian hospital: Assessing attributable
521 cost and length of stay. J Hosp Infect. 2013;85(1):54–9.

522 8. ARHAI Scotland. Scottish One Health Antimicrobial Use and Antimicrobial Resistance in
523 2020. ARHAI Scotland; 2021.

524 9. Zhou X, Willems RJL, Friedrich AW, Rossen JWA, Bathoorn E. Enterococcus faecium: from
525 microbiological insights to practical recommendations for infection control and diagnostics.
526 Antimicrob Resist Infect Control. 2020 Aug 10;9(1):130.

527 10. Quainoo S, Coolen JPM, van Hijum SAFT, Huynen MA, Melchers WJG, van Schaik W, et
528 al. Whole-Genome Sequencing of Bacterial Pathogens: the Future of Nosocomial Outbreak
529 Analysis. Clin Microbiol Rev. 2017 Oct 30;30(4):1015–63.

530 11. Pinholt M, Larner-Svensson H, Littauer P, Moser CE, Pedersen M, Lemming LE, et al.
531 Multiple hospital outbreaks of vanA Enterococcus faecium in Denmark, 2012-13,
532 investigated by WGS, MLST and PFGE. J Antimicrob Chemother. 2015;70:1–9.

533 12. Croucher NJ, Didelot X. The application of genomics to tracing bacterial pathogen
534 transmission. Curr Opin Microbiol. 2015 Feb 1;23:62–7.

535 13. Bryant JM, Grogono DM, Greaves D, Foweraker J, Roddick I, Inns T, et al. Whole-genome
536 sequencing to identify transmission of Mycobacterium abscessus between patients with
537 cystic fibrosis: a retrospective cohort study. The Lancet. 2013;381(9877):1551–60.

538 14. Worby CJ, Lipsitch M, Hanage WP. Within-Host Bacterial Diversity Hinders Accurate
539 Reconstruction of Transmission Networks from Genomic Distance Data. PLoS Comput Biol.
540 2014;10(3):e1003549.

541 15. Snitkin ES, Zelazny AM, Thomas PJ, Stock F, NISC COMPARATIVE SEQUENCING
542 PROGRAM, Henderson DK, et al. Tracking a Hospital Outbreak of Carbapenem-Resistant
543 Klebsiella pneumoniae with Whole-Genome Sequencing. Sci Transl Med.
544 2012;4(148):148ra116.

545 16. Hall MD, Holden MT, Srisomang P, Mahavanakul W, Wuthiekanun V, Limmathurotsakul D,
546 et al. Improved characterisation of MRSA transmission using within-host bacterial sequence
547 diversity. eLife. 2019;8:e46402.

548 17. Harris SR, Cartwright EJ, Török ME, Holden MT, Brown NM, Ogilvy-Stuart AL, et al. Whole-
549 genome sequencing for analysis of an outbreak of meticillin-resistant Staphylococcus
550 aureus: a descriptive study. Lancet Infect Dis. 2013;13(2):130–6.

551 18. Okoro CK, Kingsley RA, Quail MA, Kankwatira AM, Feasey NA, Parkhill J, et al. High-
552 Resolution Single Nucleotide Polymorphism Analysis Distinguishes Recrudescence and
553 Reinfection in Recurrent Invasive Nontyphoidal Salmonella Typhimurium Disease. Clin
554 Infect Dis. 2012;54(7):955–63.

555 19. Paterson GK, Harrison EM, Murray GGR, Welch JJ, Warland JH, Holden MTG, et al.
556 Capturing the cloud of diversity reveals complexity and heterogeneity of MRSA carriage,
557 infection and transmission. Nat Commun. 2015;6(6560).

558 20. Moradigaravand D, Gouliouris T, Blane B, Naydenova P, Ludden C, Crawley C, et al.
559 Within-host evolution of Enterococcus faecium during longitudinal carriage and transition to
560 bloodstream infection in immunocompromised patients. Genome Med. 2017;9(1):119.

561 21. Raven KE, Gouliouris T, Parkhill J, Peacock SJ. Genome-Based Analysis of Enterococcus
562 faecium Bacteremia Associated with Recurrent and Mixed-Strain Infection. Carroll KC,
563 editor. J Clin Microbiol. 2018;56(3):e01520-17.

564 22. Dubin KA, Mathur D, McKenney PT, Taylor BP, Littmann ER, Peled JU, et al. Diversification
565 and evolution of vancomycin-resistant Enterococcus faecium during intestinal domination.
566 Infect Immun. 2019;87:e00102-19.

567 23. Bayjanov JR, Baan J, Rogers MRC, Troelstra A, Willems RJL, van Schaik W. Enterococcus
568 faecium genome dynamics during long-term asymptomatic patient gut colonization. Microb
569 Genomics. 2019;5.

570 24. Both A, Kruse F, Mirwald N, Franke G, Christner M, Huang J, et al. Population dynamics in
571 colonizing vancomycin-resistant E. faecium isolated from immunosuppressed patients. J
572 Glob Antimicrob Resist. 2022;28:267–73.

573 25. Lebreton F, Schaik W van, McGuire AM, Godfrey P, Griggs A, Mazumdar V, et al.
574 Emergence of Epidemic Multidrug-Resistant Enterococcus faecium from Animal and
575 Commensal Strains. mBio. 2013;4(4):e00534-13.

576 26. Goodhead I, Darby AC. Taking the pseudo out of pseudogenes. Curr Opin Microbiol.
577 2015;23:102–9.

578 27. Dingle KE, Elliott B, Robinson E, Griffiths D, Eyre DW, Stoesser N, et al. Evolutionary
579 History of the Clostridium difficile Pathogenicity Locus. Genome Biol Evol. 2014;6(1):36–52.

580 28. Mäklin T, Kallonen T, David S, Boinett CJ, Pascoe B, Méric G, et al. High-resolution sweep
581 metagenomics using fast probabilistic inference. Vol. 5, Wellcome Open Research. 2021.

582 29. Anyansi C, Straub TJ, Manson AL, Earl AM, Abeel T. Computational Methods for Strain-
583 Level Microbial Detection in Colony and Metagenome Sequencing Data. Front Microbiol.
584 2020;11.

585 30. Mu A, Kwong JC, Isles NS, Silva AG da, Schultz MB, Ballard SA, et al. Reconstruction of
586 the Genomes of Drug-Resistant Pathogens for Outbreak Investigation through
587 Metagenomic Sequencing. mSphere. 2019 Feb 27;4(1):e00529-18.

588 31. Gouliouris T, Coll F, Ludden C, Blane B, Raven KE, Naydenova P, et al. Quantifying
589 acquisition and transmission of Enterococcus faecium using genomic surveillance. Nat
590 Microbiol. 2021 Jan;6(1):103–11.

591 32. Brodrick HJ, Raven KE, Harrison EM, Blane B, Reuter S, Török ME, et al. Whole-genome
592 sequencing reveals transmission of vancomycin-resistant Enterococcus faecium in a
593 healthcare network. Genome Med. 2016;8(1):4.

594 33. Raven KE, Gouliouris T, Brodrick H, Coll F, Brown NM, Reynolds R, et al. Complex Routes
595 of Nosocomial Vancomycin-Resistant Enterococcus faecium Transmission Revealed by
596 Genome Sequencing. Clin Infect Dis. 2017;64(7):886–93.

597 34. Xanthopoulou K, Wille J, Zweigner J, Lucaßen K, Wille T, Seifert H, et al. Characterization
598 of a vancomycin-resistant Enterococcus faecium isolate and a vancomycin-susceptible E.
599 faecium isolate from the same blood culture. J Antimicrob Chemother. 2020;76(4):883–6.

600 35. Cárdenas AM, Andreacchio KA, Edelstein PH. Prevalence and detection of mixed-
601 population enterococcal bacteremia. J Clin Microbiol. 2014;52(7):2604–8.

602 36. Boumasmoud M, Dengler Haunreiter V, Schweizer TA, Meyer L, Chakrakodi B, Schreiber
603 PW, et al. Genomic Surveillance of Vancomycin-Resistant Enterococcus faecium Reveals
604 Spread of a Linear Plasmid Conferring a Nutrient Utilization Advantage. mBio.
605 2022;13(2):e03771-21.

606 37. Pinholt M, Bayliss SC, Gumpert H, Worning P, Jensen VVS, Pedersen M, et al. WGS of
607 1058 Enterococcus faecium from Copenhagen, Denmark, reveals rapid clonal expansion of
608 vancomycin-resistant clone ST80 combined with widespread dissemination of a vanA-
609 containing plasmid and acquisition of a heterogeneous accessory genome. J Antimicrob
610 Chemother. 2019;74(7):1776–85.

611 38. Hashimoto Y, Taniguchi M, Uesaka K, Nomura T, Hirakawa H, Tanimoto K, et al. Novel
612 Multidrug-Resistant Enterococcal Mobile Linear Plasmid pELF1 Encoding vanA and vanM
613 Gene Clusters From a Japanese Vancomycin-Resistant Enterococci Isolate. Front
614 Microbiol. 2019;10.

615 39. Hashimoto Y, Kita I, Suzuki M, Hirakawa H, Ohtaki H, Tomita H. First Report of the Local
616 Spread of Vancomycin-Resistant Enterococci Ascribed to the Interspecies Transmission of
617 a vanA Gene Cluster-Carrying Linear Plasmid. mSphere. 2020;5(2):e00102-20.

618 40. Fujiya Y, Harada T, Sugawara Y, Akeda Y, Yasuda M, Masumi A, et al. Transmission
619 dynamics of a linear vanA-plasmid during a nosocomial multiclonal outbreak of vancomycin-
620 resistant enterococci in a non-endemic area, Japan. Sci Rep. 2021;11(1):14780.

621 41. Lee AS, White E, Monahan LG, Jensen SO, Chan R, van Hal SJ. Defining the Role of the
622 Environment in the Emergence and Persistence of vanA Vancomycin-Resistant
623 Enterococcus (VRE) in an Intensive Care Unit: A Molecular Epidemiological Study. Infect
624 Control Hosp Epidemiol. 2018;1–8.

625 42. Ford CD, Lopansri BK, Gazdik MA, Webb B, Snow GL, Hoda D, et al. Room contamination,
626 patient colonization pressure, and the risk of vancomycin-resistant Enterococcus
627 colonization on a unit dedicated to the treatment of hematologic malignancies and
628 hematopoietic stem cell transplantation. Am J Infect Control. 2016;44(10):1110–5.

629 43. McDermott H, Skally M, O'Rourke J, Humphreys H, Fitzgerald-Hughes D. Near-patient
630 environmental contamination of an intensive care unit with Vancomycin-resistant
631 Enterococci (VRE) and Extended-Spectrum Beta-Lactamase–Producing
632 Enterobacteriaceae (ESBL-E) before and after the introduction of chlorhexidine bathing for
633 patients. Infect Control Hosp Epidemiol. 2018;1–2.

634 44. Bonten MJM, Hayden MK, Nathan C, Van Voorhis J, Matushek M, Slaughter S, et al.
635 Epidemiology of colonisation of patients and environment with vancomycin-resistant
636 enterococci. Lancet. 1996;348(9042):1615–9.

637 45. Gouliouris T, Blane B, Brodrick HJ, Raven KE, Ambridge KE, Kidney AD, et al. Comparison
638 of two chromogenic media for the detection of vancomycin-resistant enterococcal carriage
639 by nursing home residents. Diagn Microbiol Infect Dis. 2016;85(4):409–12.

640 46. D'Agata EMC, Gautam S, Green WK, Tang YW. High rate of false-negative results of the
641 rectal swab culture method in detection of gastrointestinal colonization with vancomycin-
642 resistant enterococci. Clin Infect Dis. 2002;34(2):167–72.

643 47. Linfield RY, Campeau S, Injean P, Gregson A, Kaldas F, Rubin Z, et al. Practical methods
644 for effective vancomycin-resistant enterococci (VRE) surveillance: experience in a liver
645 transplant surgical intensive care unit. Infect Control Hosp Epidemiol. 2018;39(10):1178–82.

646 48. Gordon LG, Elliott TM, Forde B, Mitchell B, Russo PL, Paterson DL, et al. Budget impact
647 analysis of routinely using whole-genomic sequencing of six multidrug-resistant bacterial
648 pathogens in Queensland, Australia. BMJ Open. 2021;11(2):e041968.

649 49. Torok E, Brodrick H, Khokhar F, Blane B, Polgarova P, Brown J, et al. Prospective
650 Surveillance and Rapid Whole-Genome Sequencing Detects Two Unsuspected Outbreaks
651 of Carbapenemase-Producing Klebsiella pneumoniae in a UK Teaching Hospital. Open
652 Forum Infect Dis. 2017;4(Suppl 1):S43–4.

653 50. Sherry NL, Gorrie CL, Kwong JC, Higgs C, Stuart RL, Marshall C, et al. Multi-site
654 implementation of whole genome sequencing for hospital infection control: A prospective
655 genomic epidemiological analysis. Lancet Reg Health - West Pac. 2022;23:100446.

656 51. Forde BM, Bergh H, Cuddihy T, Hajkowicz K, Hurst T, Playford EG, et al. Clinical
657 implementation of routine whole-genome sequencing for hospital infection control of multi-
658 drug resistant pathogens [Internet]. medRxiv; 2022 [cited 2022 Aug 23]. p.
659 2022.05.02.22273921. Available from:
660 https://www.medrxiv.org/content/10.1101/2022.05.02.22273921v1

661 52. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing
662 and processing long-read sequencing data. Bioinformatics. 2018;34(15):2666–9.

663 53. Wick RR, Judd LM, Cerdeira LT, Hawkey J, Méric G, Vezina B, et al. Trycycler: consensus
664 long-read assemblies for bacterial genomes. Genome Biol. 2021;22(1):266.

665 54. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using
666 repeat graphs. Nat Biotechnol. 2019;37(5):540–6.

667 55. Ruan J, Li H. Fast and accurate long-read assembly with wtdbg2. Nat Methods.
668 2020;17(2):155–8.

669 56. Vaser R, Šikić M. Time- and memory-efficient genome assembly with Raven. Nat Comput
670 Sci. 2021;1(5):332–6.

671 57. Li H. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences.
672 Bioinformatics. 2016;32(14):2103–10.

673 58. Bruce J Walker, Thomas Abeel, Terrance Shea, Margaret Priest, Amr Abouelliel, Sharadha
674 Sakthikumar, et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection
675 and Genome Assembly Improvement | PLOS ONE. PLOS ONE. 2014;9(11):e112963.

676 59. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing
677 genome assembly and annotation completeness with single-copy orthologs. Bioinformatics.
678 2015 Oct 1;31(19):3210–2.

679 60. Seemann T. Prokka: Rapid prokaryotic genome annotation. Bioinformatics.
680 2014;30(14):2068–9.

681 61. Bortolaia V, Kaas Rs, Ruppe E, Roberts Mc, Schwarz S, Cattoir V, et al. ResFinder 4.0 for
682 predictions of phenotypes from genotypes. J Antimicrob Chemother. 2020;75(12).

683 62. Zankari E, Allesøe R, Joensen KG, Cavaco LM, Lund O, Aarestrup FM. PointFinder: a novel
684 web tool for WGS-based detection of antimicrobial resistance associated with chromosomal
685 point mutations in bacterial pathogens. J Antimicrob Chemother. 2017;72(10):2764–8.

686 63. Joensen Kg, Scheutz F, Lund O, Hasman H, Kaas Rs, Nielsen Em, et al. Real-time whole-
687 genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic
688 Escherichia coli. J Clin Microbiol. 2014;52(5).

689 64. Carattoli A, Zankari E, Garciá-Fernández A, Larsen MV, Lund O, Villa L, et al. In Silico
690 detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence
691 typing. Antimicrob Agents Chemother. 2014;58(7):3895–903.

692 65. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster
693 version of the PHAST phage search tool. Nucleic Acids Res. 2016;1–6.

694 66. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid
695 phylogenetic analysis of large samples of recombinant bacterial whole genome sequences
696 using Gubbins. Nucleic Acids Res. 2015;43(3):e15.

697 67. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data.
698 Bioinforma Oxf Engl. 2014;30(15):2114–20.

699 68. Inouye M, Dashnow H, Raven LA, Schultz MB, Pope BJ, Tomita T, et al. SRST2: Rapid
700 genomic surveillance for public health and hospital microbiology labs. Genome Med.
701 2014;6(90):1–16.

702 69. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb
703 software, the PubMLST.org website and their applications. Wellcome Open Res. 2018;3.

704 70. Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, et al. SNP-sites: Rapid
705 efficient extraction of SNPs from multi-FASTA alignments. Microb Genomics.
706 2016;2(4):e000056.

707 71. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al.
708 IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic
709 Era. Mol Biol Evol. 2020;37(5):1530–4.

710 72. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the
711 Ultrafast Bootstrap Approximation. Mol Biol Evol. 2018;35(2):518–22.

712 73. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast
713 model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587–9.

714 74. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree
715 display and annotation. Nucleic Acids Res. 2021;49(W1):W293–6.

716 75. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies
717 from short and long sequencing reads. PLoS Comput Biol. 2017;13(6):e1005595.

718 76. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2:
719 Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space.
720 Syst Biol. 2012;61(3):539–42.

721 77. Wymant C, Hall M, Ratmann O, Bonsall D, Golubchik T, de Cesare M, et al.
722 PHYLOSCANNER: Inferring Transmission from Within- and Between-Host Pathogen
723 Genetic Diversity. Mol Biol Evol. 2018;35(3):719–33.

724 78. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A
725 Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome
726 Res. 2003;13(11):2498–504.

727 79. Huebner RE, Dagan R, Porath N, Wasas AD, T M, Klugman KP. Lack of Utility of
728 Serotyping Multiple Colonies for Detection of Simultaneous Nasopharyngeal Carriage of
729 Differnt Pneumococcal Serotypes. Pediatr Infect Dis J. 2000;19(10):1017–20.

730 80. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna,
731 Austria: R Foundation for Statistical Computing; 2021. Available from: https://www.R-
732 project.org/

733 81. Wickham H. ggplot2: Elegant Graphics for Data Analysis [Internet]. Springer-Verlag New
734 York; 2016. Available from: https://ggplot2.tidyverse.org

735 82. Pedersen TL. patchwork: The Composer of Plots [Internet]. 2020. Available from:
736 https://CRAN.R-project.org/package=patchwork

737
738
739
740

741

742 **Table 1** – Characteristics of patients with rectal VREfm colonisation (n = 13)

Demographics	**N (%)**
Female	8 (61.5)
Age, median (range) years	66 (37-77)
Primary diagnosis	
Acute myeloid leukaemia	3 (23.1)
Diffuse large B cell lymphoma	3 (23.1)
Multiple myeloma	3 (23.1)
Myelodysplasia	2 (15.4)
Chronic lymphocytic leukaemia	1 (7.7)
Mantle cell lymphoma	1 (7.7)
Antimicrobial administration	
Any antibiotics in the 7d prior to positive screen	12 (92.3)
Any antibiotics in the 6m prior to positive screen	12 (92.3)
Vancomycin in the 7d prior to positive screen	1 (7.7)
Vancomycin in the 6m prior to positive screen[a]	3 (30.0)
Outcomes within 60d of VREfm positive screen	
VREfm bloodstream infection	3 (23.1)
ICU admission	1 (7.7)
Death	0 (0)

743 [a] Information available for 10 patients

744

745 **Table 2** – MLST Sequence Types Detected

Patient ID	Sample ID	Sample Type	Sample Date (days from start of study)	STs detected (n, %)	Maximum pairwise SNP distance within sample	Median (IQR) pairwise SNP distance within sample
P2	VRED01	Rectal	6	80 (14, 100)	2	0 (0 - 1)
P6	VRED16	Rectal	27	1424 (14, 100)	2	0 (0 - 1)
P7	VRED10	Rectal	18	80 (14, 100)	3	0 (0 - 1)
P9	VRED02	Rectal	6	80 (14, 100)	1	0 (0 - 0)
P9	VRED03	Rectal	8	80 (14, 100)	0	0 (0 - 0)
P9	VRED09	Rectal	17	80 (14, 100)	1	0 (0 - 0)
P9	VRED18	Blood	52	80 (1, 100)	NA	NA
P9	VRED19	Blood	52	80 (1, 100)	NA	NA
P9	VRED20	Blood	52	80 (1, 100)	NA	NA
P9	VRED21	Blood	52	80 (1, 100)	NA	NA
P14	VRED07	Rectal	17	1424 (10, 71.4); 80 (4, 26.6)	491	2 (1 - 482)
P20	VRED15	Rectal	27	80 (14, 100)	3	1 (0 - 2)
P20	VRED17	Rectal	30	80 (14, 100)	3	1 (0 - 1)
P24	VRED13	Rectal	26	1424 (14, 100)	2	0 (0 - 0)
P33	VRED12	Rectal	24	80 (14, 100)	2	0 (0 - 1)
P34	VRED04	Rectal	9	1424 (14, 100)	1	0 (0 - 0)
P34	VRED05	Rectal	14	1424 (14, 100)	4	0 (0 - 1)
P49	VRED06	Rectal	15	80 (10, 71.4); 1424 (3, 21.4); 789 (1, 7.1)	490	0 (0 - 482)
P49	VRED08	Rectal	17	1424 (14, 100)	2	0 (0 - 1)
P49	VRED14	Blood	30	1424 (1, 100)	NA	NA
P50	VRED11	Rectal	23	1424 (13, 92.9); 1659 (1, 7.1)	4429	1 (0 - 2)

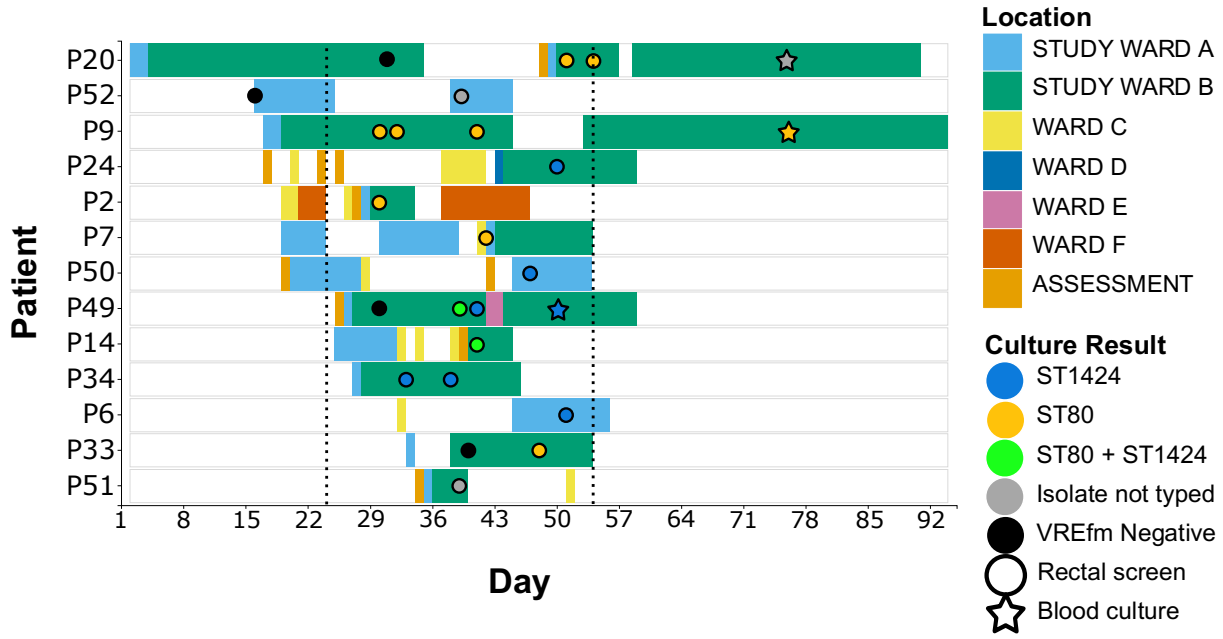
746
747

748 **Table 3** – Presence of AMR Genes

Gene	Phenotypic resistance	ST80, n = 130		ST1424, n = 97		All Isolates, n = 229	
		n (%)	Genetic element	n (%)	Genetic element	n (%)	Summary
aac(6')-aph(2'')	Amikacin, Gentamicin, Kanamycin, Streptomycin, Tobramycin	65 (50.0)	p1_VRED06-10	94 (96.9)	p1_VRED06-02	160 (69.9)	Variable in ST80/ST1424/ST789
aac(6')-Ii	Gentamicin, Tobramycin	130 (100)	Chromosome	97 (100)	Chromosome	229 (100)	All isolates
ant(6)-Ia	Streptomycin	130 (100)	p3_VRED06-10	0 (0)	NA	131 (57.2)	All ST80/ST789
ant(9)-Ia	Spectinomycin	0 (0)	NA	97 (100)	Chromosome	97 (42.4)	All ST1424
aph(3')-III	Amikacin, Kanamycin, Neomycin	130 (100)	p3_VRED06-10	97 (100)	p2_VRED06-02	228 (99.6)	All ST80/ST789/ST1424
dfrG	Trimethoprim	0 (0)	NA	97 (100)	Chromosome	97 (42.4)	All ST1424
erm(A)	Clindamycin, Erythromycin, Quinupristin	0 (0)	NA	97 (100)	Chromosome	97 (42.4)	All ST1424
erm(B)	Clindamycin, Erythromycin, Quinupristin	129 (99.2)	p1_VRED06-10, p3_VRED06-10	97 (100)	p2_VRED06-02	228 (99.6)	All except one ST80 isolate
msr(C)	Erythromycin, Quinupristin	130 (100)	Chromosome	97 (100)	Chromosome	229 (100)	All isolates
tet(L)	Doxycycline, Tetracycline	0 (0)	NA	0 (0)	NA	1 (0.4)	Only ST1659
tet(M)	Doxycycline, Minocycline, Tetracycline	0 (0)	NA	60 (61.9)	Chromosome	61 (26.6)	Variable in ST1424/ST1659
tet(S)	Doxycycline, Minocycline, Tetracycline	130 (100)	p3_VRED06-10	0 (0)	NA	131 (57.2)	All ST80/ST789
VanA	Teicoplanin, Vancomycin	97 (100)	p2_VRED06-02	97 (100)	p2_VRED06-02	229 (100)	All isolates

749
750

751



752

753 **Figure 1** – Patient timeline. Each row denotes the location of a patient during admission, blocks

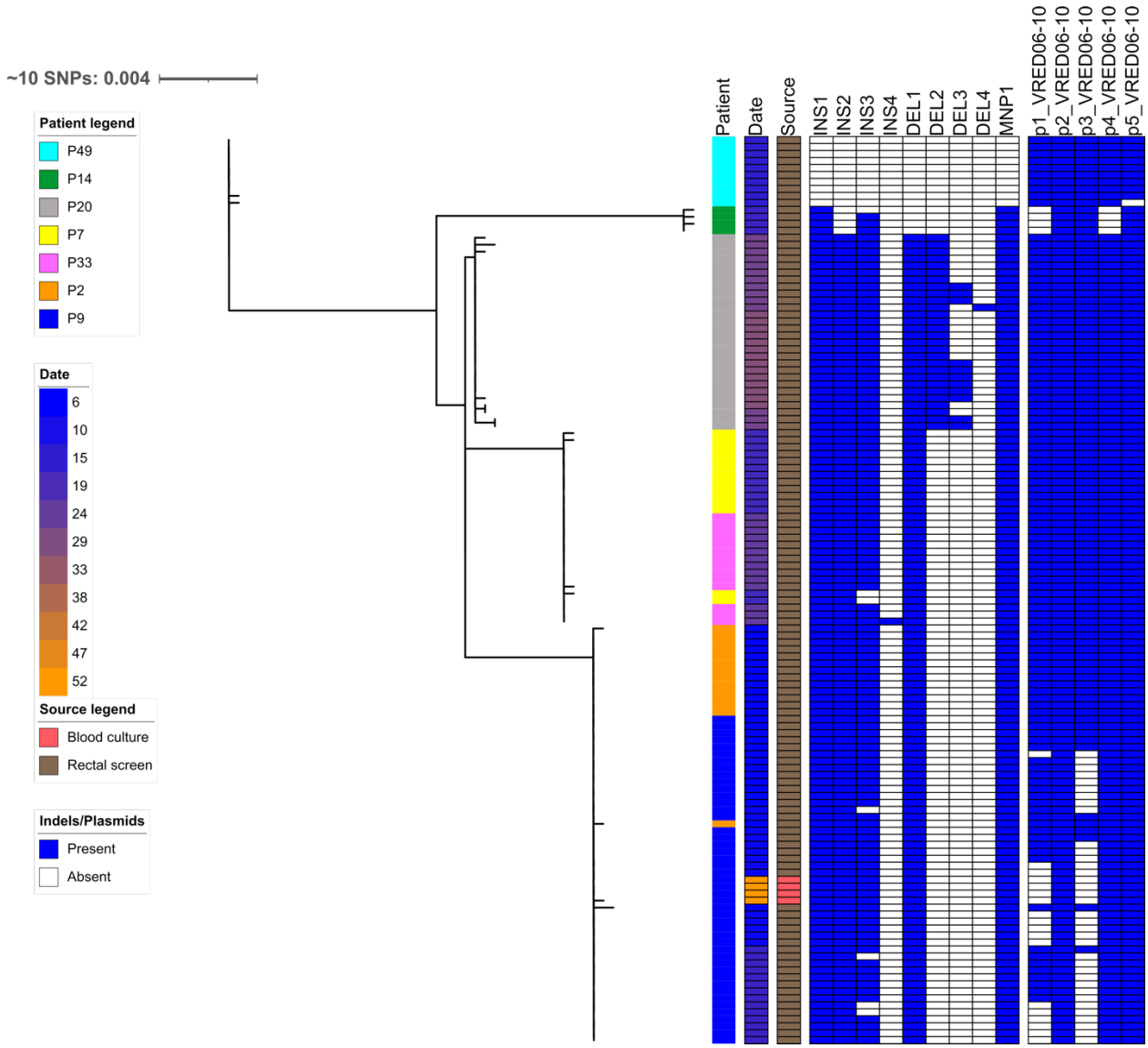
754 denote hospital stay, circles denote VREfm cultures, stars denote bloodstream isolates, dotted

755 lines indicate the start and end of prospective collection of screening isolates for this study. This

756 study was undertaken mainly within Wards A and B, although patients were moved to different

757 wards within the hospital during their stay and were often admitted through the assessment unit.

758



759

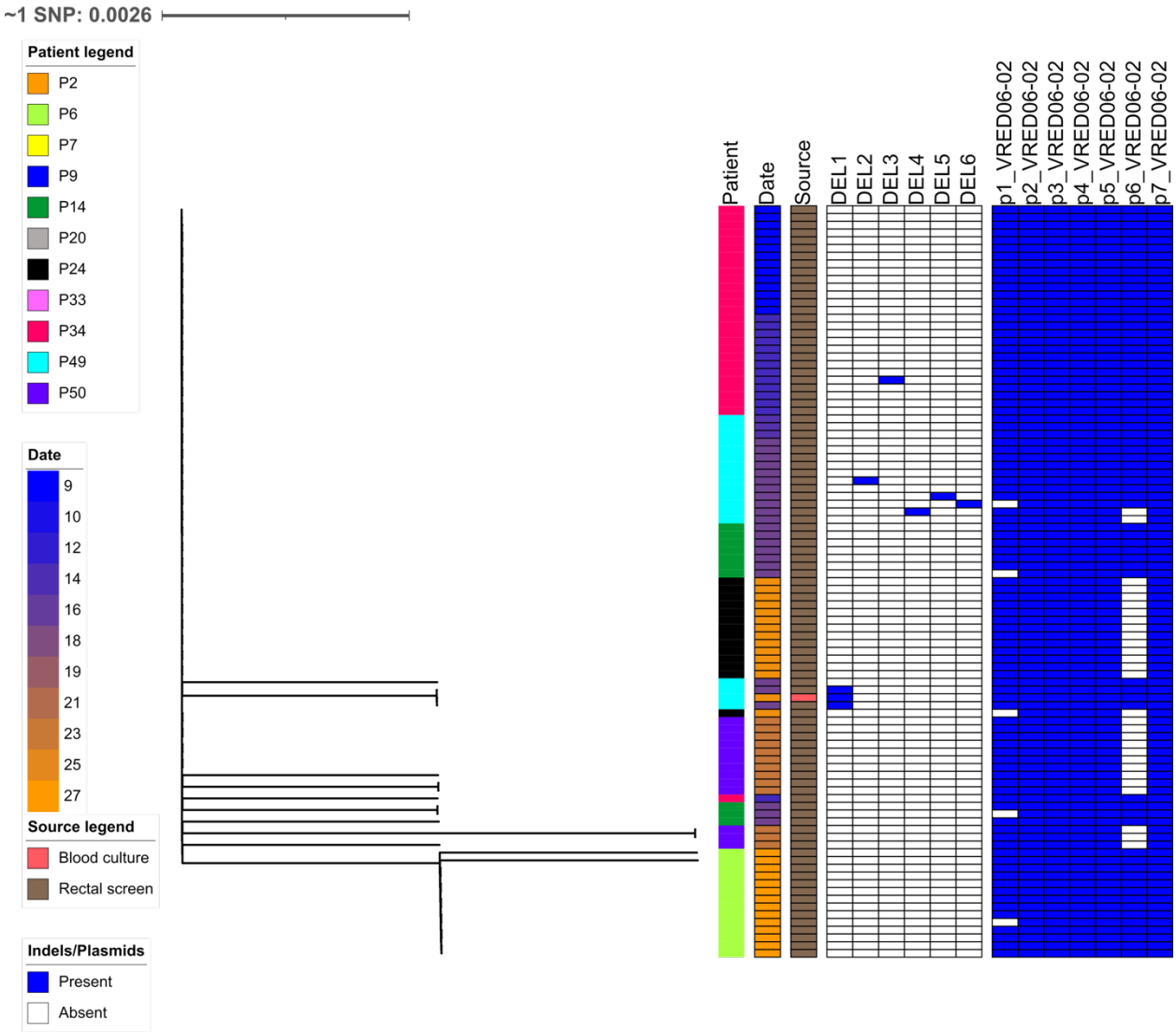
760 **Figure 2** – Phylogeny of ST80 isolates showing structured population with three patient specific

761 clusters and two clusters indicating putative patient-patient transmission of VREfm. All ST80

762 isolates mapped to VRED06-10 chromosome and phylogeny built on SNP sites (n=96) after

763 removal of putative transposable and recombinant regions.

764



765

766 **Figure 3** – Phylogeny of ST1424 isolates showing homogeneous population suggestive of

767 recent transmission outbreak. ST1424 isolates (n=98) mapped to VRED06-02 chromosome and

768 phylogeny built on SNP sites (n=13) after removal of putative transposable and recombinant

769 regions.

770

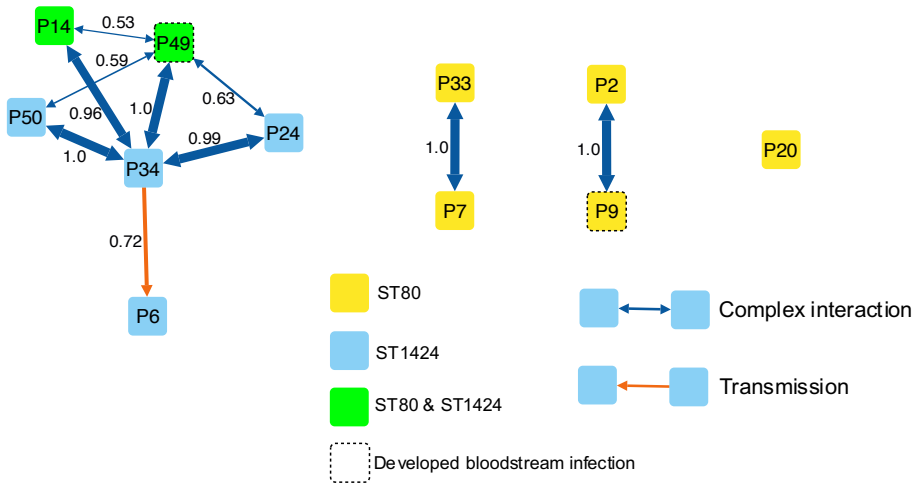
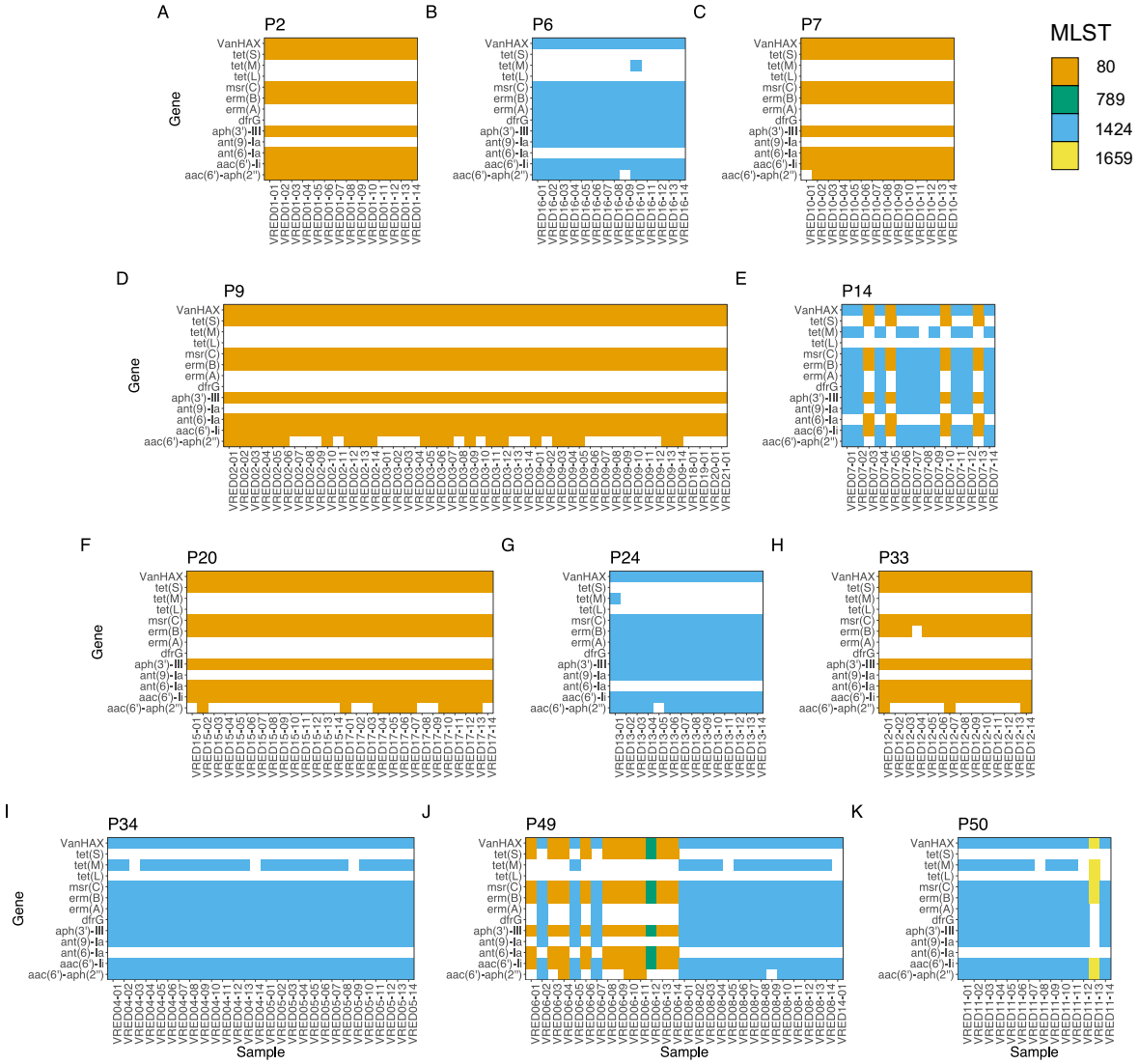


Figure 4 – Phyloscanner transmission network. Each patient is represented by a node coloured by detection of the two outbreak STs. Edge thickness corresponds to fraction of Phyloscanner trees with given relationship, relationship fraction is printed alongside each edge, and edge colour based on type of relationship (orange, direct transmission; blue, transmission but direction unclear).



779

780 **Figure 5** – Detection of antimicrobial resistance genes. Panels represent different patients;

781 resistance genes are plotted on the y-axis and isolates on the x-axis. Presence of a gene is

782 represented by a filled square and coloured based on the MLST ST of the isolate.

783