View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by University of Bath Research Portal

Citation for published version:
Pinholt, M, Bayliss, SC, Gumpert, H, Worning, P, Jensen, VVS, Pedersen, M, Feil, EJ & Westh, H 2019, 'WGS of 1058 *Enterococcus faecium* from Copenhagen, Denmark, reveals rapid clonal expansion of vancomycin-resistant clone ST80 combined with widespread dissemination of a vanA-containing plasmid and acquisition of a heterogeneous accessory genome', *Journal of Antimicrobial Chemotherapy*, vol. 74, no. 7, pp. 1776-1785. https://doi.org/10.1093/jac/dkz118

DOI:
10.1093/jac/dkz118

Publication date:
2019

Document Version
Peer reviewed version

Link to publication

This is a pre-copyedited, author-produced version of an article accepted for publication in Journal of Antimicrobial Chemotherapy following peer review. The version of record Mette Pinholt, Sion C Bayliss, Heidi Gumpert, Peder Worning, Veronika V S Jensen, Michael Pedersen, Edward J Feil, Henrik Westh, WGS of 1058 Enterococcus faecium from Copenhagen, Denmark, reveals rapid clonal expansion of vancomycin-resistant clone ST80 combined with widespread dissemination of a vanA-containing plasmid and acquisition of a heterogeneous accessory genome, Journal of Antimicrobial Chemotherapy, Volume 74, Issue 7, July 2019, Pages 1776–1785,is available online at: https://doi.org/10.1093/jac/dkz118

University of Bath

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 22. Sep. 2020

Whole genome sequencing of 1058 *Enterococcus faecium* from Copenhagen, Denmark, reveals rapid clonal expansion of vancomycin-resistant clone ST80 combined with widespread dissemination of a *vanA*-containing plasmid and acquisition of a heterogeneous accessory genome

Mette PINHOLT[1,2*#], Sion C. BAYLISS[3#], Heidi GUMPERT[1], Peder WORNING[1], Veronika Vorobieva S. JENSEN[4], Michael PEDERSEN[5], Edward J. FEIL[3], Henrik WESTH[1,2].

1. Department of Clinical Microbiology, Hvidovre University Hospital, Hvidovre, Denmark.
2. Institute of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark.
3. Milner Centre for Evolution, Department of Biology and Biochemistry, University of Bath, Bath, United Kingdom.
4. Department of Clinical Microbiology, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark.
5. Department of Clinical Microbiology, Herlev University Hospital, Herlev, Denmark.

* Corresponding author: Mette Pinholt, Department of Clinical Microbiology, Copenhagen University Hospital, Hvidovre Hospital, Hvidovre, Denmark. Tel: +45 38623210; Fax: +45 38623357; E-mail: mette.pinholt@regionH.dk

# Contributed equally

Running title: VREfm expansion and acquisition of accessory genes

Abstract

Objectives: From 2012-2015, a sudden huge increase of vancomycin-resistant (*vanA*) *Enterococcus faecium* (VREfm) was observed in the Capital Region of Denmark. Clonal relatedness of VREfm and vancomycin-susceptible *E. faecium* (VSEfm) was investigated, transmission events between hospitals were identified and the pan-genome and plasmids from the largest VREfm clonal group were characterized.

Methods: WGS of 1058 *E. faecium* isolates was carried out on the Illumina platform to perform SNP-analysis and to identify the pan-genome. One isolate was also sequenced on the PacBio platform to close the genome. Epidemiological data were collected from laboratory information systems.

Results: Phylogeny of 892 VREfm and 166 VSEfm revealed a polyclonal structure with a single clonal group (ST80) accounting for 40% of the VREfm isolates. VREfm and VSEfm co-occurred within many clonal groups; however, no VSEfm were related to the dominant VREfm group. A similar *vanA* plasmid was identified in ≥99% of isolates belonging to the dominant group and 69% of the remaining VREfm. Ten plasmids were identified in the completed genome and approximately 29% of this genome consisted of dispensable accessory genes. The size of the pan-genome among isolates in the dominant group was 5,905 genes.

Conclusions: Most likely, VREfm emerged due to import of a successful VREfm clone which has rapidly transmitted to the majority of hospitals in the region whilst simultaneously disseminating a *vanA* plasmid to pre-existing VSEfm. Acquisition of a heterogeneous accessory genome may account for the success of this clone by facilitating adaption to new environmental challenges.

Introduction

Enterococcus faecium, a commensal in the gut, has emerged as an increasingly important nosocomial pathogen causing bacteraemia, intra-abdominal, urinary tract and intravenous -catheter related infections.[1] Acquired resistance to ampicillin, gentamicin (high-level) and vancomycin has increased worldwide among hospital-associated *E. faecium* narrowing treatment options of enterococcal infections.[2-5] Patients, asymptomatically colonized with vancomycin-resistant *E. faecium* (VREfm) in the GI-tract act as both a reservoir and a source for dissemination of VREfm into the hospital environment.[6] This makes infection control of this problem pathogen highly challenging. Previous studies have identified two to ten asymptomatic VREfm carriers for each patient with a clinical VREfm isolate.[6]

Vancomycin is first line treatment of infections caused by *E. faecium*. Therefore, it was of great concern when the incidence of *vanA* VREfm started to increase in the Capital Region of Denmark in 2012.[7] More than 1,500 VREfm patients were identified from 2012 to 2015, contrasting with only 9 patients infected or colonised in 2011.[8, 9]

We recently reported on WGS of 495 VREfm isolates identified from the Capital Region of Denmark from 2012-2014 and observed a polyclonal outbreak characterized by the spread of a highly successful plasmid (pHvH-V24, KX574671.1), which conferred resistance to vancomycin through the presence of a *vanA* gene complex.[10] This *vanA* plasmid had been acquired by >90% of the VREfm clonal groups. Therefore, we developed the hypothesis that a VREfm clone, containing a *vanA* plasmid, has been introduced to the Capital Region of Denmark and the plasmid was disseminated via horizontal gene transfer to pre-existing hospital-adapted vancomycin-susceptible *E. faecium* (VSEfm) clonal groups.

Hospital-associated *E. faecium* are known to rapidly acquire mobile genetic elements (MGE) and, as a result, they harbour plasmids and insertion elements that confer resistance, virulence and other clinically relevant phenotypes.[11, 12] More detailed molecular data on MGEs, the accessory genome and comparison to VSEfm are therefore necessary to understand how VREfm and resistance determinants disseminate in the hospital environment.

Here, we applied short- and long-read sequencing along with epidemiological data to understand the sudden increase of VREfm in the low prevalence Capital Region of Denmark from 2012-2015. The aims of the study were 1) to analyse the genetic relatedness between temporally and geographically matched VREfm and VSEfm and to detect the spread of the *vanA* plasmid between different clones, 2) to identify VREfm transmission events between hospitals and 3) to characterize the pan-genome and plasmids within the largest clonal group accounting for 40% of the VREfm isolates.

Methods

Setting and bacterial isolates

The study was conducted in the Capital Region of Denmark (1.7 million inhabitants) from January 2012 to December 2015. The healthcare system is served by 11 hospitals (one tertiary hospital, four large acute care secondary hospitals and six minor secondary hospitals). VREfm isolates from all Departments of Clinical Microbiology (n=3) were submitted consecutively for WGS from 2012-March 2014 (n=338) whereas only isolates from the Southern part of the region were submitted from April 2014-2015 (n=554). VREfm screening isolates were identified using selective chromogenic medium and MALDI-TOF. VREfm in clinical samples were identified using MALDI-TOF and vancomycin disc diffusion according to EUCAST.[13] Only *vanA* VREfm were included as the vast majority of the isolates contained that genotype (five isolates contained *vanB*). One isolate per patient was included.

VSEfm blood isolates from 2013-2014 (n=113) and VSEfm non-blood isolates from 2013-2015 (n=53) were included in the study. Ampicillin resistance was determined using disc diffusion according to EUCAST.[13]

Date of sampling, submitting hospital, hospital ward, sample site and sample type (clinical or screening sample) were collected for all isolates from the laboratory information systems.

VREfm identified from 2012-2014 were sequenced in a previous study.[10] All VREfm from 2015 (n=399) and VSEfm (n=166) were sequenced for this study.

Whole genome short read sequencing and comparative analysis

BDNA extraction and sequencing using Illumina technology were conducted as previously described.[7, 10] Sequence reads were mapped to the V24 reference genome (see below) and to individual plasmids using SMALT.[14] Reads at indel sites were realigned using the GATK toolbox.[15] SNPs were called using samtools.[16] The variant call file (VCF) was filtered for variants supported by a minimum read depth of 4 (minimum 2 per strand), >30 map quality, >50 average base quality, no significant strand bias and >75% of reads supporting the variant. Indels were additionally confirmed using pindel.[17] Repeat regions of >50bp and any regions <100 bp between two repeat regions, which are notoriously problematic for short read mapping, were identified using nucmer and removed from the VCF resulting in consensus sequences.[18]

A phylogenetic tree (neighbour-joining tree) including all VREfm and VSEfm isolates was constructed using RapidNJ and tree visualization was done using FigTree.[19, 20] In addition, a subtree of isolates belonging to the largest clonal group was constructed using RAxML using the general time reversible model and gamma model of rate heterogeneity. Recombination regions were identified and removed using ClonalFrameML with a kappa setting of 4, and priors of R/theta = 0.462, 1/delta = 0.004, nu = 0.042, allowing the rate of recombination to vary per branch.[21, 22]

MLST was determined *in silico* using pubMLST.

Generating a complete genome sequence for a vanA E. faecium isolate using hybrid sequencing

A screening isolate, V24, which was an early representative of the largest VREfm clonal group, was identified as a reference strain. DNA extraction, Illumina sequencing and PacBio sequencing of V24 were performed as described previously.[7, 10] We achieved a total of 241,971 PacBio reads with a mean length of 5.8 kb.

V24 was assembled using a hybrid assembly methodology using both short, low error rate Illumina reads and PacBio long-reads. Reads were assembled by SPAdes v3.6.2 using the --careful, --pacbio, and a manual range of k-mer sizes of 21, 33, 55, 67, 77, 87, 97, 107, 117 and 127.[23] Contigs were scaffolded using SSPACE-longRead,[24] and gap filling was performed using PBJelly.[25] A single chromosome and ten non-chromosomal contigs/scaffolds >600 bp were assembled. The chromosome was found to be largely contiguous with previously published reference genomes after alignment using MAUVE.[26-28] The ten non-chromosomal contig/scaffold sequences were searched against the NCBI nr/nt database using default megablast settings. Contigs and scaffolds were circularized using an in-house script as described previously.[10] All but one contig were circularisable using this methodology. PCRs were performed to circularize the last contig, but despite repeated attempts, this did not succeed.

Characterization of the complete genome V24

The chromosome and all plasmids were annotated using RAST.[29] Resistance genes, virulence genes, prophages and plasmid replicons were identified via Antibiotic Resistance Gene-ANNOTation, virulenceFinder, Phast and PlasmidFinder online tools, respectively.[30-32] Plasmids were searched for Toxin-antitoxin systems (axe-txe and ω-ε-ζ) via BLAST according to Rosvoll *et al*.[33] Finally, copy number of plasmids was determined as the mean depth coverage (excluding repeat regions) of the plasmid divided by the mean depth coverage of the chromosome.

Presence of ten plasmids in VREfm and VSEfm isolates

Illumina reads from all VREfm and VSEfm isolates were mapped to the ten plasmid sequences generated from the reference isolate, V24. The depth of coverage, breadth of coverage and the total number of SNPs (excluding repeat regions) was determined. A plasmid was characterized as present in a sample when at least 85% of the reference sequence was covered by sufficient reads to call a reference or variant base and the number of SNPs was ≤5 SNPs per 1000 bp (~99.5% similarity). A reduced breadth of coverage of ≥80% was accepted for plasmid pV24-2 due to large amounts of repeat regions which led to lower breadth of read coverage (86%) even when mapping the short reads used to create the reference genome, V24, against itself.

Pan-genome analysis

Reads were assembled *de novo* using Velvet v1.0.11 and contigs were annotated by Prokka v1.11. The pan-genome of the largest clonal group was identified by Roary using default settings.[34, 35] Core genes were defined as those present in 99% of the isolates with a minimum percentage amino-acid identity of 95%. We excluded 27 out of 361 isolates (7.5%) from the pan-genome analysis due to poor assemblies (assemblies of ≥800 contigs).

Ethics

The project has been approved by the Danish Data Protection Agency (2012-58-0004/AHH-2015-047) and the Danish Health and Medicines Authority (3-3013-1118/1).

Results

Descriptive data

In total, 1,058 *E. faecium* isolates (892 *vanA* VREfm and 166 ampicillin resistant VSEfm) were included in the study. A third (n=292, 33%) of the VREfm isolates were from clinical samples and of these isolates 12% (n=35) were from blood cultures. The vast majority of the VREfm isolates (n=832, 93%) were identified in the five largest hospitals in the region and most VSEfm isolates (n=156, 94%) were identified in the same five hospitals. More epidemiological data are presented in Table S1.

Population structure of VREfm

SNP-based phylogeny revealed high diversity between hospital-associated VREfm isolates (Figure 1A). The isolates were separated into 16 main groups of ≥3 isolates, 10 minor groups of 2 isolates and 25 singletons. The majority, n=728 (81%) of the isolates belong to five groups consisting of 47-361 VREfm isolates. Clonal group 2_ST80 was the dominant clone from 2012-2015 (Figure 1B). Excluding group 2_ST80, the dominant clonal groups were 1_ST117 and 3_ST192 in 2012-2013 until they were replaced by clonal groups 8_ST203 and 13_ST80 in 2014-2015.

Comparative analysis of temporally and geographically matched VREfm and VSEfm

The phylogeny showed that vancomycin-resistant and -susceptible isolates co-occurred within many of the clonal groups (Figure 1A). In total, n=146 (88%) VSEfm clustered together with VREfm. However, the largest VREfm group (2_ST80, n=361), responsible for the first local outbreak of VREfm, contained no VSEfm. Also, clonal groups 8_ST203 and 13_ST80, which were prevalent from 2015, only contained 1 and 0 VSEfm, respectively.

Epidemiology and spread of clonal group 2_ST80

The largest clonal group, 2_ST80, was analysed further in order to understand the rapid dissemination throughout the region. A maximum likelihood phylogenetic tree, generated for group 2_ST80, was annotated with hospital location and sampling date (Figure 2A). Further details are presented interactively using Microreact (https://microreact.org/project/BkNqA1EDX, use Google Chrome as browser).[36] Only samples from 2012 until March 2014 were included (n=142) because sampling coverage was uniform across the whole region for that time period. Very few SNP differences (mean pairwise SNPs = 8) were observed between the isolates. The epidemic curve shows that 90% (n=9) of the isolates in 2012 were sampled from Hospital 1, suggesting that Hospital 1 was the site into which group 2_ST80 was introduced (Figure 2B). Subsequently, the clone spread to five more hospitals within six months showing that clonal group 2_ST80 became widespread in the region within less than a year. The phylogenetic tree branches out in four subgroups. One branch is almost unique to Hospital 2 and the three other branches contain isolates from four or five hospitals. Assuming that each branch corresponds to a new introduction, this suggests that clonal group 2_ST80 has been transmitted several times between hospitals in the region (frequent cross-hospital transmission).

Genome summary of V24

A representative isolate, V24, of clonal group 2_ST80, was selected for PacBio sequencing to investigate this successful group further. The genome of V24 included a single circular chromosome (2.7 Mb and 2,625 coding sequences (CDS)), nine circular plasmids with sizes ranging from 4,304 bp to 172,811 bp (4-178 CDS) and one partial or linear plasmid of 103,806 bp (133 CDS) (Table 1). In total, the genome contained 3,139 CDS and approximately 14% of the genome was located on plasmids. The best hit in a blast search against the NCBI nr/nt database for each plasmid were previously published *E. faecium* plasmids with query coverages of >60% and similarity of >90%. The exception was the linear/partial plasmid, pV24-2, that was novel, with query coverage of 15% for the best hit.

The *vanA* plasmid, pV24-5, was closed in a previous study.[10] Re-analyses of PacBio and Illumina data in this study have extended the plasmid with 4,016 bp (Figure S1). This extension was confirmed by PCR. The added sequence contained four hypothetic proteins and a replication protein. The plasmid contained several resistance genes and a BLAST search revealed that the plasmid was similarly to the non-conjugative plasmid pS177 (Table 1).[37] Compared to the Tn*1546* prototype characterized by Arthur *et al*,[38] the *vanA* transposon was truncated with a deletion of the transposase and 156 bp of the resolvase. An insertion element (IS*1251*) was inserted in the *vanSH* intergenic region. The loss of the transposase suggests that the element is no longer mobile and genes involved in the mobilization of the plasmid could not be identified.

Genes responsible for conjugation and horizontal transfer (*traG*) were identified in two other plasmids, pV24-1 and pV24-3 (Table 1),[39] and a BLAST search revealed that pV24-3 was similar to a highly conjugative plasmid pZB18.[40]

A search for prophages identified both intact and incomplete prophage regions in the chromosome and four plasmids. Importantly, the linear/partial plasmid, pV24-2, had three incomplete prophage regions of which many coding sequences corresponded to sequences in a linear bacteriophage recently described in a closed *E. faecalis* genome.[41] This may explain why pV24-2 was not circularisable.

Distribution of V24 plasmids among the E. faecium population

V24 plasmids (pV24-1- pV24-10) were much more frequent in VREfm than in VSEfm isolates (Table 2). Within group 2_ST80 to which V24 belongs, the prevalence of V24 plasmids was highly variable. Two plasmids were present in <10% of the isolates and four plasmids were present in >90% of the isolates. This pattern shows that isolates belonging to the same clonal group contain different large accessory elements. The conjugative plasmid, pV24-3, was only present in group 2_ST80 isolates identified from 2012-13. After 2013, the plasmid disappeared.

The *vanA* plasmid (pV24-5) was frequent, n=727 (81%) and spread across all but one VREfm clonal group (Group 8_ST203). The remaining V24 plasmids were also rare in group 8_ST203 indicating that there has been no exchange of plasmids between the two largest VREfm clonal groups.

Pan-genome analysis

The pan-genome of group 2_ST80 was identified to assess the distribution between core genes and accessory genes within a clonal group of highly similar isolates determined by SNP analysis (mean pairwise SNPs = 8). The pan-genome contained 5,905 genes of which 1,684 were core genes, 2,223 were soft core genes (present in ≥95%) and 2960 were cloud genes (present in <15%) (Figure 3). We know that the V24 isolate, belonging to group 2_ST80, contained 3,139 CDS. A calculation shows that approx. 71% (2,223 soft core genes/3,139 CDS) of the V24 genome consisted of conserved and essential soft core genes and the remaining 29% of the genome consisted of non-essential accessory genes.

Discussion

This study provides insights into a polyclonal VREfm outbreak which arose in the Capital Region of Denmark in 2012. The largest clonal group (2_ST80) accounting for 40% of the isolates was responsible for the first local VREfm hospital outbreak. This was followed by a rapid clonal expansion and transmission into multiple hospitals in the region. No VSEfm isolates clustered in this group despite the large number of isolates sampled, so it can be assumed that clonal group 2_ST80 was imported to the region as a vancomycin-resistant clone in 2012. A *vanA* plasmid, pV24-5, initially identified in a group

2_ST80 isolate (V24) was observed in 81% of all VREfm isolates (Table 2). We also observed that VSEfm and VREfm co-occurred in many clonal groups which supports the hypothesis that concurrently with clonal expansion, group 2_ST80 successfully disseminated pV24-5 via HGT to pre-existing hospital-adapted VSEfm and new VREfm clones were generated.

Group 2_ST80 spread rapidly (within 6-9 months) to multiple hospitals within the Capital Region after its introduction to hospital 1 (Figure 2). This remarkable fast transmission is likely a product of the highly interlinked healthcare network within the region, with medical specializations consolidated into only a few Hospitals, which results in frequent transfer of patients between hospitals (approximately 1,500 patients per month between hospitals, data from Department of Financial Affairs, the Capital Region of Denmark). VREfm colonized patients are asymptomatic carriers, and transfer of those patients represents a great risk of starting a new VREfm outbreak. In previous studies, clonally related VREfm isolates were identified in hospitals that rarely share patients and a community reservoir was suggested.[27, 28, 42-46] In contrast, our genomic analyses and epidemiological data support a hospital reservoir for the dominant clonal group (2_ST80).

From the beginning of the VREfm outbreak, barrier precautions, extra cleaning, disinfection with chlorine and VREfm screening of contact patients were initiated. Active surveillance screening programmes were not systematically introduced at the same time in all hospitals in the region which may explain continued expansion of VREfm.

The increase of VREfm and results from the present study prompted several measures to control VREfm: 1) Switch from a culture based methods to *vanA*/*vanB* PCR to achieve faster results (<24 hours) and higher sensitivity, 2) Active surveillance screenings of wards with ≥2 VREfm patients, 3) VREfm screening of patients transferred from another hospital to high risk departments, 4) Screening and isolation of VREfm patients readmitted within six months and 5) real-time sequencing of VREfm isolates (results within 1-2 weeks) to include/exclude patients from outbreaks and to better understand transmission routes.

In 2015, clonal group 8_ST203 emerged (Figure 1). This group did not contain the successful *vanA* plasmid, pV24-5, and only one VSEfm clustered in the group suggesting import and spread of yet another successful VREfm clonal group. Recently, the Danish Reference Laboratory at Statens Serum Institut closed a genome of an early (October 2014) group 8_ST203 isolate.[42, 47] A *vanA* plasmid, pVRE1589 (CP020486), different from pV24-5 (Figure S1), was identified. Ninety-seven percent of the isolates in clonal group 8_ST203 contained pVRE1589 (Henrik Hasman, Department of Bacteria, Parasites and Fungi, Statens Serum Institut, personal communication). Introduction of a clonal group containing a new *vanA* plasmid indicates that the hospital environment in the Capital Region has become favorable for VREfm within the last five years, as the prevalence of VREfm was less than 20 isolates per year before 2012.[8, 9] Important factors to drive the current emergence of VREfm could be

increasing use of vancomycin and metronidazole due to a large *Clostridium difficile* outbreak, shift of empirical therapy from cefuroxime to piperacillin-tazobactam and increasing tolerance for disinfectants among *E. faecium* isolates.[48-52]

WGS-based typing is an important tool for infection control to define nosocomial outbreaks, include/exclude patients from an outbreak and to identify transmission events. In this work we identified widespread dissemination of a *vanA*-containing plasmid. Short read sequencing allowed for cost-effective epidemiological analysis and isolate screening on a large scale. Long-read sequencing allowed us to resolve problematic regions of the genome, such as plasmids, which were relevant to the current work. Based on this study we recommend using a combination of long-read and short-read sequencing for bacteria with highly variable accessory genome content to interpret outbreaks correct.

The present study identified a large number of MGEs in VREfm with 14% of the coding capacity located on plasmids in V24. This was in line with the pan-genome analysis of group 2_ST80 which identified acquisition of a highly heterogeneous accessory genome (Figure 3) suggesting frequent flux of MGEs. This flexible and mobile genome may contribute to the success of the clone by providing it with ready access to genes which enable it to adapt to new environmental changes. Lack of the CRISPR-Cas system, which provides bacteria with sequence-specific, acquired defense against plasmids and phage, was suggested as an explanation for the high number of MGEs in enterococci.[53] Previous studies have reported up to six plasmids per *E. faecium* genome with 9-10% of the genome located on plasmids.[27, 28, 42-44] Potential donors to the heterogeneous accessory genome could be other intestinal bacteria as *E. faecium* is part of a complex community in the GI-tract. However, previous studies showed that sequence importation in relation to recombination was mainly from other *E. faecium* populations.[54, 55]

The *vanA* plasmid, pV24-5, was highly similar to plasmid pS177, which was identified in a VREfm isolate from USA more than ten years ago and recently a pS177-like plasmid was reported in Australia, suggesting that the plasmid is globally spread.[37, 44] The plasmid, pS177, is a non-conjugative plasmid lacking the transposase in Tn*1546* which indicates that the element is not mobile.[37] Presumably, the *vanA* plasmid, pV24-5, used the conjugation machinery from another conjugative plasmid to disseminate via HGT. Adding PacBio sequencing to close the genome of V24 revealed a conjugative plasmid, pV24-3, which was present in all group 2_ST80 isolates identified from 2012 until September 2013 (Table 2). This suggests that pV24-3 may have played an important role in the horizontal transfer of the *vanA* plasmid. From 2014, pV24-3 disappeared from the clone and was either replaced by another conjugative plasmid or the ability to disseminate the *vanA* plasmid was lost. This may explain why clonal groups 1_ST117 and 3_ST192 almost disappeared in 2015 (Figure 1). However, the *vanA* plasmid, pV24-5 was also present in clonal group 13_ST80 which became quiet dominant in 2015. Our knowledge of conjugative transfer systems and plasmid interactions is sparse and hopefully, long-read sequencing will

improve our understanding of this. Also, experimental characterization of the required attributes for mobility and other accessory genetic elements contributing to this would be valuable.

The strength of this study is the large number of contemporary VREfm and VSEfm isolates that are included. Inclusion of VSEfm data, which is new compared to our previous study on VREfm has provided valuable information on the close relationship between VREfm and VSEfm.[10] Completion of a *vanA* genome enabled us to assess the plasmid content and provided us with a reference genome closely related to the outbreak we are investigating. However, this study has limitations. We expect that the plasmid population is much larger than investigated in this study. In addition, mixed infections may have been overlooked due to the sampling methodology of selecting only one colony for sequencing from each patient.

In conclusion, we have shown that a successful VREfm clone was imported to a VREfm low prevalence region in Denmark in 2012. Both clonal expansion and generation of new VREfm clonal groups via horizontal transfer of a *vanA* plasmid to pre-existing VSEfm contributed to the sudden increase and persistence of VREfm. This combination of clonal spread and plasmid spread must be recognized to interpret VREfm hospital outbreaks correctly and highlights the added value of long-read sequencing in WGS.

The acquisition of a highly heterogeneous accessory genome and the high number of plasmids within group 2_ST80 showed that there is no single successful combination of accessory genes, suggesting that VREfm isolates belonging to this clone can rapidly acquire genes to optimize survival in the harsh hospital environment between transmissions.

Acknowledgements

Data has been presented at the 25th European Congress of Clinical Microbiology and Infectious Diseases, Copenhagen, Denmark, 2015 (oral presentation O168) and at the 11th International Meeting on Microbial Epidemiological Markers, Estoril, Portugal, 2016 (oral presentation OP43). Sequence data from 2012-2014 was presented in Journal of Antimicrobial chemotherapy, 2017; 72: 40-7. We thank Louise Christensen, Susanne Rhode, Marianne S. Studstrup, Casper Dam-Nielsen and Maria Kristin Bjornsdóttir for excellent technical assistance. We thank the SSAC Foundation, The A.P. Møller foundation (Fonden til Laegevidenskabens Fremme) and the Danish Ministry of Health.

Funding

This work was supported by the SSAC Foundation, the Danish Ministry of Health, The A.P. Møller foundation (Fonden til Laegevidenskabens Fremme) and internal funding.

Transparency declarations

None to declare.

Accession numbers

Illumina sequence data for all isolates (VREfm and VSEfm) were deposited in the ENA under the projects PRJEB14625 and PRJEB28731. The closed genome sequences of V24 were deposited at Genbank under accession numbers CP036151-CP036161.

Reference List

1. Pinholt M, Ostergaard C, Arpi M, *et al*. Incidence, clinical characteristics and 30-day mortality of enterococcal bacteraemia in Denmark 2006-2009: a population-based cohort study. *Clin Microbiol Infect* 2014; **20**: 145-51.

2. Antimicrobial resistance interactive database: EARS-Net. http://ecdc.europa.eu/en/healthtopics/antimicrobial_resistance/database/Pages/graph_reports.aspx.

3. Surveillance report. Antimicrobial resistance surveillance in Europe 2014. http://ecdc.europa.eu/en/publications/Publications/antimicrobial-resistance-europe-2014.pdf.

4. The Australian Group on Antimicrobial Resistance: http://www.agargroup.org/.

5. News From Centers for Disease Control and Prevention. Mixed News on Hospital Infections. *JAMA* 2016;315:1694 .

6. Cattoir V, Leclercq R. Twenty-five years of shared life with vancomycin-resistant enterococci: is it time to divorce? *J Antimicrob Chemother* 2013; **68**: 731-42.

7. Pinholt M, Larner-Svensson H, Littauer P, *et al*. Multiple hospital outbreaks of vanA Enterococcus faecium in Denmark, 2012-13, investigated by WGS, MLST and PFGE. *J Antimicrob Chemother* 2015; **70**: 2474-82.

8. Surveillance data from the Capital Region of Denmark: www.regionh.dk/hospitalsinfektioner/overvågningsdata/Sider/default.aspx.

9. Surveillance data VRE Denmark: https://statistik.ssi.dk//sygdomsdata#!/?sygdomskode=VRE&xaxis=Aar&show=Graph&datatype=Laboratory

10. Pinholt M, Gumpert H, Bayliss S, *et al*. Genomic analysis of 495 vancomycin-resistant Enterococcus faecium reveals broad dissemination of a vanA plasmid in more than 19 clones from Copenhagen, Denmark. *J Antimicrob Chemother* 2017; **72**: 40-7.

11. Guzman Prieto AM, van SW, Rogers MR, *et al*. Global Emergence and Dissemination of Enterococci as Nosocomial Pathogens: Attack of the Clones? *Front Microbiol* 2016; **7**: 788.

12. Wurster JI, Saavedra JT, Gilmore MS. Impact of Antibiotic Use on the Evolution of Enterococcus faecium. *J Infect Dis* 2016; **213**: 1862-5.

13. EUCAST: Clinical breakpoints: http://www.eucast.org/clinical_breakpoints/.

14. SMALT: Pairwise Sequence Alignment Program. Welcome Trust Sanger Institute, Hinxton, UK. http://www.sanger.ac.uk/science/tools/smalt-0.

15. GATK: https://software.broadinstitute.org/gatk/.

16. Li H, Handsaker B, Wysoker A, *et al*. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; **25**: 2078-9.

17. Ye K, Schulz MH, Long Q, *et al.* Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* 2009; **25**: 2865-71.

18. Kurtz S, Phillippy A, Delcher AL, *et al.* Versatile and open software for comparing large genomes. *Genome Biol* 2004; **5**: R12.

19. FigTree. A graphic viewer of phylogenetic trees. http://tree.bio.ed.ac.uk/software/fiftree v1.4.0.

20. Simonsen M, Mailund T, Pedersen CNS. Rapid Neighbour Joining. In *Proceedings of the 8th Workshop in Algorithms in Bioinformatics (WABI),* LNBI 5251, 113-122, Springer Verlag, October 2008. doi:10.1007/978-3-540-87361-7_10.

21. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014; **30**: 1312-3.

22. Didelot X, Wilson DJ. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLoS Comput Biol* 2015; **11**: e1004041.

23. Bankevich A, Nurk S, Antipov D, *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012; **19**: 455-77.

24. Boetzer M, Pirovano W. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics* 2014; **15**: 211.

25. English AC, Salerno WJ, Reid JG. PBHoney: identifying genomic variants via long-read discordance and interrupted mapping. *BMC Bioinformatics* 2014; **15**: 180.

26. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 2010; **5**: e11147.

27. Lam MM, Seemann T, Bulach DM, *et al.* Comparative analysis of the first complete Enterococcus faecium genome. *J Bacteriol* 2012; **194**: 2334-41.

28. Qin X, Galloway-Pena JR, Sillanpaa J, *et al.* Complete genome sequence of Enterococcus faecium strain TX16 and comparative genomic analysis of Enterococcus faecium genomes. *BMC Microbiol* 2012; **12**: 135.

29. Aziz RK, Bartels D, Best AA, *et al.* The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 2008; **9**: 75.

30. Carattoli A, Zankari E, Garcia-Fernandez A, *et al.* In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 2014; **58**: 3895-903.

31. Zhou Y, Liang Y, Lynch KH, *et al.* PHAST: a fast phage search tool. *Nucleic Acids Res* 2011; **39**: W347-W352.

32. Antibiotic Resistance Gene-ANNOTation: https://omictools.com/arg-annot-tool.

33. Rosvoll TC, Pedersen T, Sletvold H, *et al.* PCR-based plasmid typing in Enterococcus faecium strains reveals widely distributed pRE25-, pRUM-, pI. *FEMS Immunol Med Microbiol* 2010; **58**: 254-68.

34. Page AJ, Cummins CA, Hunt M, *et al.* Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 2015; **31**: 3691-3.

35. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014; **30**: 2068-9.

36. Argimon S, Abudahab K, Goater RJ, *et al.* Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *Microb Genom* 2016; **2**: e000093.

37. Halvorsen EM, Williams JJ, Bhimani AJ, *et al.* Txe, an endoribonuclease of the enterococcal Axe-Txe toxin-antitoxin system, cleaves mRNA and inhibits protein synthesis. *Microbiology* 2011; **157**: 387-97.

38. Arthur M, Molinas C, Depardieu F, *et al.* Characterization of Tn1546, a Tn3-related transposon conferring glycopeptide resistance by synthesis of depsipeptide peptidoglycan precursors in Enterococcus faecium BM4147. *J Bacteriol* 1993; **175**: 117-27.

39. Arends K, Celik EK, Probst I, *et al.* TraG encoded by the pIP501 type IV secretion system is a two-domain peptidoglycan-degrading enzyme essential for conjugative transfer. *J Bacteriol* 2013; **195**: 4436-44.

40. Zheng B, Tomita H, Xiao YH, *et al.* Molecular characterization of vancomycin-resistant enterococcus faecium isolates from mainland China. *J Clin Microbiol* 2007; **45**: 2813-8.

41. Brede DA, Snipen LG, Ussery DW, *et al.* Complete genome sequence of the commensal Enterococcus faecalis 62, isolated from a healthy Norwegian infant. *J Bacteriol* 2011; **193**: 2377-8.

42. Hoffmann M, Yao K, Allard M, *et al.* Complete Genome Sequence of a Vancomycin-Resistant Sequence Type 203 Enterococcus faecium Strain with vanA Belonging to Complex Type 859. *Microbiol Resour Announc* 2018; **7**.

43. Lam MM, Seemann T, Tobias NJ, *et al.* Comparative analysis of the complete genome of an epidemic hospital sequence type 203 clone of vancomycin-resistant Enterococcus faecium. *BMC Genomics* 2013; **14**: 595.

44. McKenney PT, Ling L, Wang G, *et al.* Complete Genome Sequence of Enterococcus faecium ATCC 700221. *Genome Announc* 2016; **4**.

45. van Hal SJ, Beukers AG, Timms VJ, *et al.* Relentless spread and adaptation of non-typeable vanA vancomycin-resistant Enterococcus faecium: a genome-wide investigation. *J Antimicrob Chemother* 2018; **73**: 1487-91.

46. Raven KE, Reuter S, Reynolds R, *et al.* A decade of genomic history for healthcare-associated Enterococcus faecium in the United Kingdom and Ireland. *Genome Res* 2016; **26**: 1388-96.

47. Hammerum AM, Baig S, Kamel Y, *et al.* Emergence of vanA Enterococcus faecium in Denmark, 2005-15. *J Antimicrob Chemother* 2017; **72**: 2184-90.

48. The weekly newsletter on surveillance and outbreaks of infectious diseases. https://www.ssi.dk/aktuelt/nyhedsbreve/epi-nyt/2015/uge-10---2015.

49. DANMAP 2016. Use of antimicrobial agents and occurrence of antimicrobial resistance in bacteria from food animals, food and humans in Denmark.https://www.danmap.org/Downloads/Reports.aspx.

50. Alotaibi SMI, Ayibiekea A, Pedersen AF, *et al.* Susceptibility of vancomycin-resistant and -sensitive Enterococcus faecium obtained from Danish hospitals to benzalkonium chloride, chlorhexidine and hydrogen peroxide biocides. *J Med Microbiol* 2017; **66**: 1744-51.

51. Pidot SJ, Gao W, Buultjens AH, *et al.* Increasing tolerance of hospital Enterococcus faecium to handwash alcohols. *Sci Transl Med* 2018; **10**.

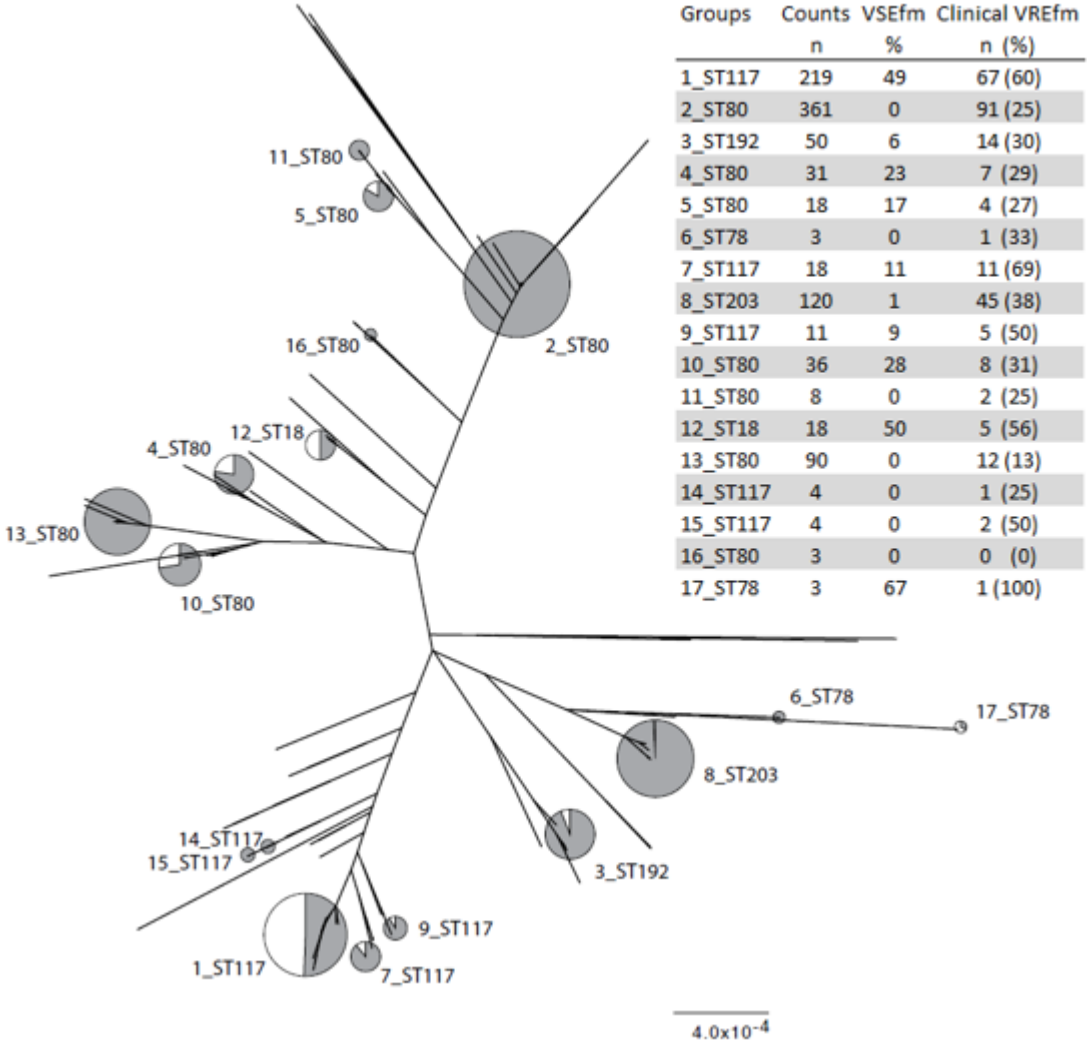
52. Gouliouris T, Warne B, Cartwright EJP, *et al.* Duration of exposure to multiple antibiotics is associated with increased risk of VRE bacteraemia: a nested case-control study. *J Antimicrob Chemother* 2018; **73**: 1692-9.

53. Palmer KL, Gilmore MS. Multidrug-resistant enterococci lack CRISPR-cas. *MBio* 2010; **1**.

54. van Hal SJ, Ip CL, Ansari MA, *et al.* Evolutionary dynamics of Enterococcus faecium reveals complex genomic relationships between isolates with independent emergence of vancomycin resistance. *Microb Genom* 2016; **2**.

55. de BM, van SW, Cheng L, *et al.* Recent recombination events in the core genome are associated with adaptive evolution in Enterococcus faecium. *Genome Biol Evol* 2013; **5**: 1524-35.

A



Groups	Counts	VSEfm	Clinical VREfm
	n	%	n (%)
1_ST117	219	49	67 (60)
2_ST80	361	0	91 (25)
3_ST192	50	6	14 (30)
4_ST80	31	23	7 (29)
5_ST80	18	17	4 (27)
6_ST78	3	0	1 (33)
7_ST117	18	11	11 (69)
8_ST203	120	1	45 (38)
9_ST117	11	9	5 (50)
10_ST80	36	28	8 (31)
11_ST80	8	0	2 (25)
12_ST18	18	50	5 (56)
13_ST80	90	0	12 (13)
14_ST117	4	0	1 (25)
15_ST117	4	0	2 (50)
16_ST80	3	0	0 (0)
17_ST78	3	67	1 (100)

B

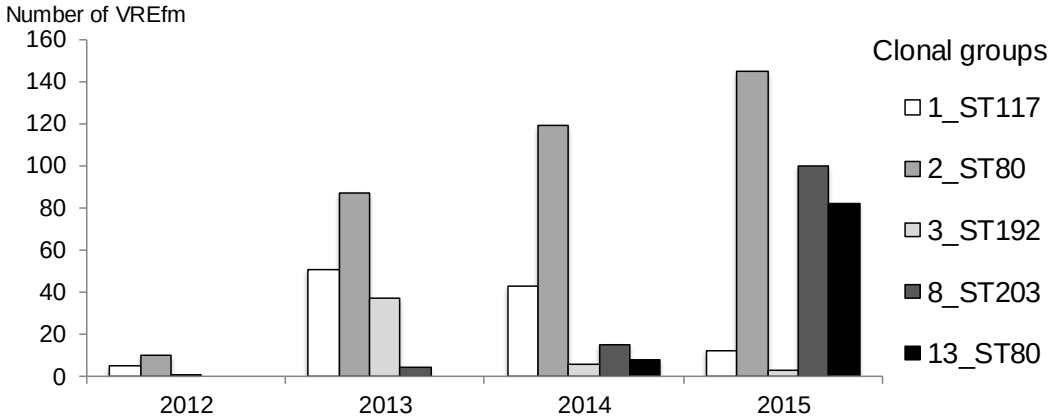


Figure 1: A. SNP-based phylogeny of 892 *vanA E. faecium* (VREfm) and 166 vancomycin-susceptible *E. faecium* (VSEfm) identified in the Capital Region of Denmark from 2012-2015. Eight outliers have been removed from the phylogenetic tree. Each circle represents a main group and was scaled in relation to the number of isolates in the group. The pie charts show the proportion of VREfm (grey) and VSEfm (white) isolates within each main group. Circles are located near the centre of the group on the phylogenetic tree.

The group numbers are assigned according to when the group was first identified and according to ST. Group identification number, number of isolates, percentage of VSEfm isolates in each main group and percentage of clinical VREfm out of all VREfm are given in the table. B. The number of VREfm belonging to the five largest clonal groups divided by calendar years.

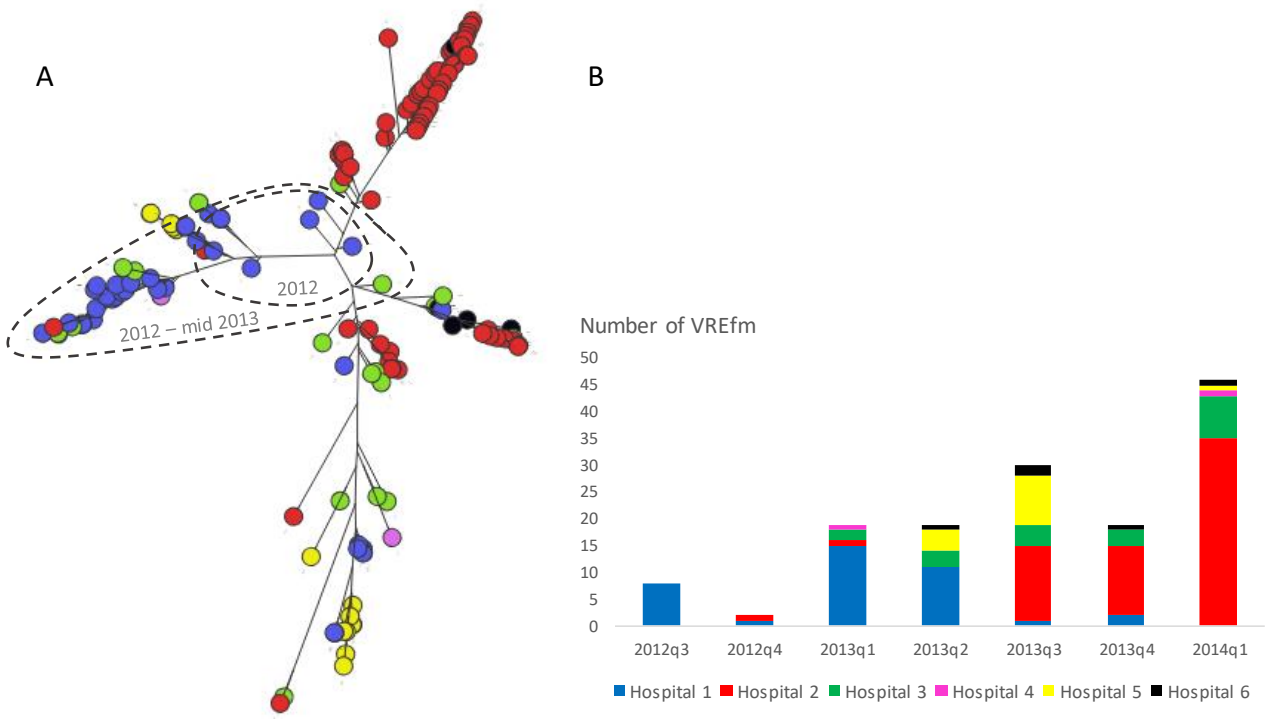


Figure 2: A. Phylogenetic tree of clonal group 2_ST80, 2012- March 2014 with masking of recombination. Coloured circles represent the hospital origin of each of the isolates. The nested circular shapes represent the approx. sampling times with inner circle contains isolates sampled until December 2012 and the outer circle contains isolates sampled until June 2013. Isolates outside the circles are sampled from July 2013 to March 2014. Data are presented interactively at https://microreact.org/project/BkNqA1EDX (use Google Chrome as browser).[36] B. Epidemic curve of clonal group 2_ST80, 2012- March 2014. This figure appears in colour in the online version of *JAC* and in black and white in the printed version of *JAC*.

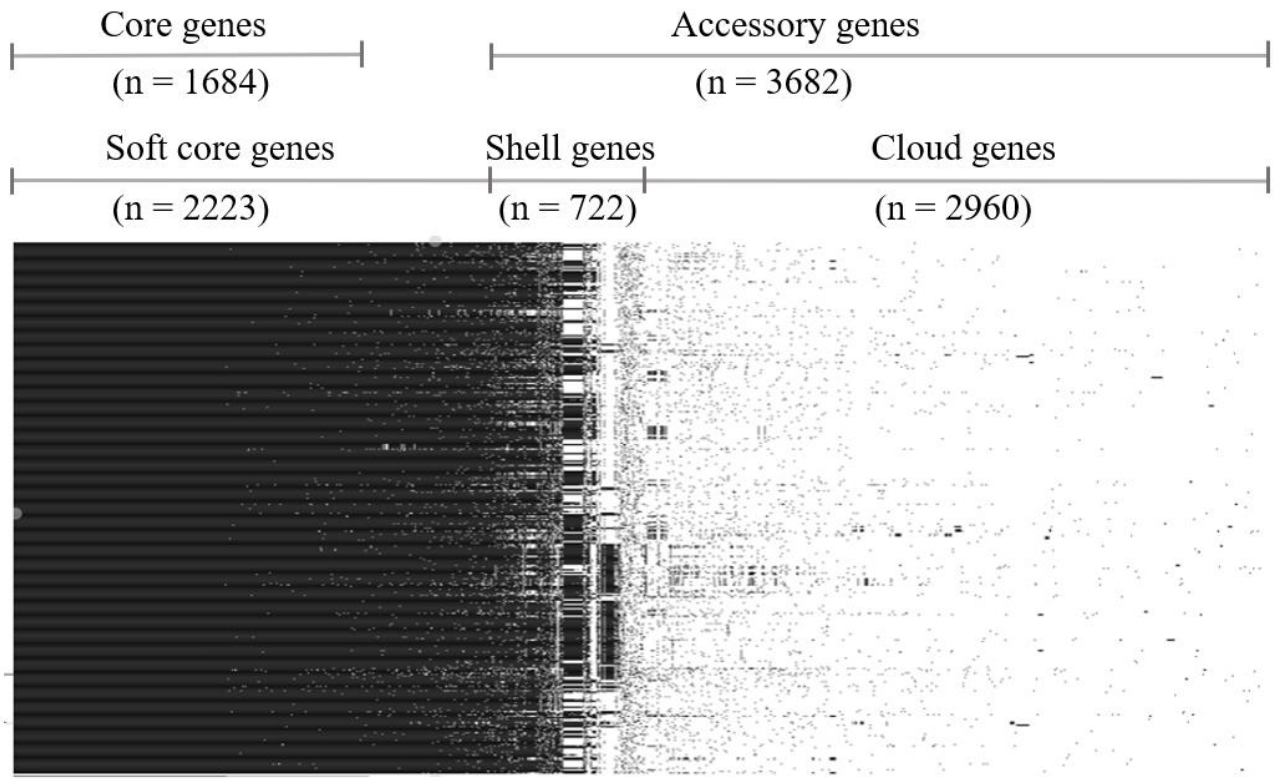


Figure 3: Visualization of the pan-genome of group 2_ST80 (n=334 isolates). Each line represents an isolate. Presence of a gene is indicated in black and absence of a gene in white. Core genes (present in ≥99% of the isolates), soft core genes (present in ≥95% of the isolates), accessory genes (present in <95% of the isolates), shell genes (present in <95% and ≥15% of the isolates) and cloud genes (present in <15% of the isolates) are indicated. The isolates are vertically ordered based on similarity in a NJ-tree.

Genome	Size (bp)	GC content (%)	CDS	Circula-risable	Copy number	Virulence genes	Resistance Genes	*rep* types	Prophage regions	Other features	Blast search in NCBI
Chromo-some	2,720,495	38.2	2.625	Yes	1	*Acm*	*msrC, dfrG, tetM*		2 intact (28.7 kb, 35.9 kb), 1 questionable (14.8 kb) and 1 incomplete (18 kb)		
pV24-1	172,811	35.3	178	Yes	1	*HylEfm*	*Aac6_Aph2*	15	1 intact (24.8 kb)	*traG*	Ef AUS00233, coverage 73%, 99% identical
pV24-2	103,806	33.4	133	No	3				3 incomplete (7.9 kb, 10.8 kb, 11.6 kb)		Ef 6E6 p1, coverage 15%, 99% identical
pV24-3	61,681	33.0	70	Yes	2					*traG*	Ef pZB18 DNA, coverage 90%, 99% identical
pV24-4	41,792	35.4	47	Yes	3		*aph(3')-III, ant(6)-Ia, sat4A, tetS*	2	1 intact (24.5 kb)		Ef p5753cB, partial sequence, coverage 67%, 99% identical
pV24-5	41,388	35.4	51	Yes	3		*aph(3')-III, ant(6)-Ia, sat4A, ermB, vanA*	17	1 incomplete (24.5 kb)	*Txe-axe*	Ef pS177, coverage 88%, 99% identical
pV24-6	9,317	30.8	12	Yes	3					Bacte-riocin	Ef AUS0085 p4, coverage 100%, 100% identical
pV24-7	6,173	35.5	8	Yes	6			11		Bacte-riocin	Ef pB82,coverage 100%, 100% identical
pV24-8	5,954	33.3	6	Yes	7			18			Ef pVEF4,coverage 61%, 91% identical
pV24-9	4,465	32.1	4	Yes	5						Ef pHY DNA, coverage 61%, 99% identical
pV24-10	4,304	37.1	5	Yes	10						Ef AUS0004 p2, coverage 82%, 97% identical

Table 1: Summary of the *vanA E. faecium* genome of V24.
CDS: coding sequences, *Acm*: collagen adhesion gene, *HylEfm*: Hyaluronidase gene, *Aac6_Aph2*, *aph(3')-III*, *ant(6)-Ia* and *sat4A*: resistance genes of aminoglycosides, *vanA*: resistance genotype of vancomycin, *msrC* and *ermB*: resistance genes of lincosamides, *dfrG*: resistance gene of trimethoprim, *tetS* and *tetM*: resistance genes of tetracyclin, *traG*: transfer gene, Txe-axe: Toxin-antitoxin system, Ef: *E. faecium.*

Group	n	pV24-1	pV24-2	pV24-3	pV24-4	pV24-5	pV24-6	pV24-7	pV24-8	pV24-9	pV24-10
		---------- n (%) ----------									
1_ST117	111	4 (4)	98 (88)	0 (0)	1 (1)	99 (89)	5 (5)	108 (97)	107 (96)	29 (26)	21 (19)
2_ST80	361	15 (4)	227 (63)	103 (29)	167 (46)	357 (99)	29 (8)	359 (99)	350 (97)	340 (94)	257 (71)
3_ST192	47	43 (91)	12 (26)	0 (0)	0 (0)	44 (94)	1 (2)	42 (89)	42 (89)	45 (96)	42 (89)
4_ST80	24	2 (8)	12 (50)	0 (0)	1 (4)	19 (79)	0 (0)	15 (63)	17 (71)	19 (79)	15 (63)
5_ST80	15	0 (0)	13 (87)	0 (0)	0 (0)	15 (100)	0 (0)	14 (93)	4 (27)	2 (13)	3 (20)
7_ST117	16	0 (0)	1 (6)	0 (0)	1 (6)	15 (94)	0 (0)	16 (100)	16 (100)	2 (13)	2 (13)
8_ST203	119	1 (1)	37 (31)	0 (0)	1 (1)	1 (1)	0 (0)	118 (99)	1 (1)	8 (7)	4 (3)
9_ST117	10	10 (100)	9 (90)	0 (0)	0 (0)	7 (70)	0 (0)	10 (100)	10 (100)	3 (30)	3 (30)
10_ST80	26	0 (0)	11 (42)	0 (0)	0 (0)	20 (77)	0 (0)	26 (100)	13 (50)	26 (100)	13 (50)
11_ST80	8	0 (0)	8 (100)	0 (0)	0 (0)	8 (100)	0 (0)	8 (100)	8 (100)	0 (0)	0 (0)
12_ST18	9	2 (22)	5 (56)	0 (0)	0 (0)	4 (44)	0 (0)	3 (33)	3 (33)	2 (22)	1 (11)
13_ST80	90	84 (93)	46 (51)	0 (0)	1 (1)	89 (99)	0 (0)	90 (100)	89 (99)	90 (100)	61 (68)
Others	56	20 (36)	33 (59)	2 (4)	3 (5)	45 (80)	8 (14)	46 (82)	44 (79)	34 (61)	23 (41)
Total **VRE**fm	892	181 (20)	512 (57)	105 (12)	175 (20)	723 (81)	43 (5)	854 (96)	704 (79)	600 (67)	444 (50)
Total **VSE**fm	166	16 (10)	33 (20)	0 (0)	0 (0)	0 (0)	12 (7)	123 (74)	116 (70)	31 (19)	14 (8)

Table 2: Plasmid distribution (pV24-1 to pV24-10) among vancomycin-resistant *E. faecium* (VREfm), vancomycin-susceptible *E. faecium* (VSEfm) and the 12 largest VREfm clonal groups.