Infection, Genetics and Evolution 54 (2017) 74–80

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research paper

Time to review the gold standard for genotyping vancomycin-resistant enterococci in epidemiology: Comparing whole-genome sequencing with PFGE and MLST in three suspected outbreaks in Sweden during 2013–2015

Birgitta Lytsy [a], Lars Engstrand [a,b,c], Åke Gustafsson [a], Rene Kaden [a,*]

[a] Uppsala University, Department of Medical Sciences, Uppsala, Sweden
[b] Karolinska Institute Solna, Sweden
[c] Science for Life Laboratory Solna, Sweden

ARTICLE INFO

Article history:
Received 17 March 2017
Received in revised form 9 June 2017
Accepted 12 June 2017
Available online 15 June 2017

Keywords:
PFGE
MLST
NGS
ANI
VRE
Cut off point WGS

ABSTRACT

Vancomycin-resistant enterococci (VRE) are a challenge to the health-care system regarding transmission rate and treatment of infections. VRE outbreaks have to be controlled from the first cases which means that appropriate and sensitive genotyping methods are needed.

The aim of this study was to investigate the applicability of whole genome sequencing based analysis compared to Pulsed-Field Gel Electrophoresis (PFGE) and Multi-Locus Sequence Typing (MLST) in epidemiological investigations as well as the development of a user friendly method for daily laboratory use.

Out of 14,000 VRE - screening samples, a total of 60 isolates positive for either *vanA* or *vanB* gene were isolated of which 38 were from patients with epidemiological links from three suspected outbreaks at Uppsala University Hospital. The isolates were genotypically characterised with PFGE, MLST, and WGS based core genome Average Nucleotide Identity analysis (cgANI). PFGE was compared to WGS and MLST regarding reliability, resolution, and applicability capacity.

The PFGE analysis of the 38 isolates confirmed the epidemiological investigation that three outbreaks had occurred but gave an unclear picture for the largest cluster. The WGS analysis could clearly distinguish six ANI clusters for those 38 isolates.

As result of the comparison of the investigated methods, we recommend WGS-ANI analysis for epidemiological issues with VRE. The recommended threshold for *Enterococcus faecium* VRE outbreak strain delineation with core genome based ANI is 98.5%.

All referred sequences of this study are available from the NCBI BioProject number PRJNA301929.

1. Introduction

Since 2008, Sweden has experienced four major nosocomial outbreaks of VRE of *vanB* genotype of which the largest outbreak occurred between 2013 and 2014 in Gavle County with over 300 patients involved. PFGE was the molecular method used at the Public Health Agency in Sweden for genotyping the isolates. PFGE is a stable and reproducible method and considered the "gold standard" for genotyping VRE in nosocomial outbreaks (Valdezate et al., 2009; Werner et al., 2012). PFGE is time consuming and highly qualified and experienced laboratory staff is needed for data evaluation (Tenover et al., 1995; van Belkum, 1994). The method is based on restriction of the whole bacterial genome followed by scoring the obtained size of DNA fragments (Werner, 2013). Standardisations for inter-laboratory comparisons do not exist for typing of VRE –isolates (Cookson et al., 2007). Thus, the method is applicable to compare isolates for regional surveillance in which the isolates have to be compared in one laboratory. MLST is the standard method for epidemiological investigations for large scale international comparisons (Maiden et al., 1998). MLST is not as discriminating as PFGE, however the sequence types (ST) are defined and can be exchanged between laboratories worldwide (Ruiz-Garbajosa et al., 2006). Whole Genome Sequencing (WGS) could be an alternative in the molecular epidemiological investigation of VRE (Kao et al., 2014). In addition to providing the same genetic data as MLST, many other genetic loci can be used in single nucleotide polymorphism (SNP) analysis or Core Genome MLST (cgMLST) (de Been et al., 2015).

The genetic distance between two whole genomes can be calculated by the average nucleotide identity (ANI). Results of ANI analysis correlates strongly with DNA – DNA hybridization. A value of 70% in DNA –

* Corresponding author.
E-mail address: rene.kaden@akademiska.se (R. Kaden).

http://dx.doi.org/10.1016/j.meegid.2017.06.010

DNA reassociation corresponds to 93–94% in ANI analysis and the majority of bacterial strains with an ANI >94% belongs to the same species (Konstantinidis and Tiedje, 2005). Due to the high resolution of WGS, even strains of the same species can be discriminated. WGS is therefore a suitable tool for molecular epidemiologic analysis in outbreak investigations. The increasing number of available WGS data makes it possible to assign new outbreak related genomes to existing data.

The purpose of this study was to compare PFGE and MLST with WGS- ANI regarding reliability, discriminatory power, epidemiological concordance and convenience criteria such as software based analysis, availability of databases, and comparability of the results of different laboratories for epidemiological molecular typing during outbreak investigations involving VRE-isolates. We furthermore aimed to develop an easy to use WGS-ANI workflow and to determine the cut off criteria for outbreak isolate assignment to use in a clinical microbiology laboratory setting.

2. Material and methods

2.1. Epidemiological investigation

According to the national recommendation of the Public Health Agency in Sweden, an epidemiological investigation should be carried out whenever VRE is isolated in a clinical culture from a patient admitted to a hospital or a nursing home in order to detect outbreaks at an early stage. The patient should be isolated in a single room with an en-suite bathroom and maximal contact precautions should be undertaken to prevent transmission. Active surveillance samples should be undertaken repeatedly in order to find all cases. Every patient admitted to the same ward as a patient with VRE should be screened for VRE in faeces, wounds and urine. Screening for VRE should be done once weekly and when patients are discharged from the ward for as long as there is a known VRE-positive patient present in the ward.

The infection prevention and control (IPC) team of Uppsala University Hospital (UUH) leads the epidemiological investigation in Uppsala County and recommends interventions for staff in the wards in order to prevent transmission. To investigate epidemiological links, a locally developed software for daily tracing of patients and their movements in the hospital wards and out-patient clinics were used for this study. Contacts found were sampled from faeces, wounds and urine according to the national policy.

2.2. Whole genome sequencing (WGS)

All screening samples from contacts, from active surveillance, and all clinical cultures collected in health-care settings in the county of Uppsala were sent for microbiological diagnostics to the clinical microbiological laboratory of UUH. All VRE isolates that were related to the outbreaks in 2013–2015 were cultured on Haematin agar plates and incubated overnight at 37 °C. Pure colonies were transferred to Brain Heart Infusion with a Vancomycin disc (5 μg; Oxoid) and incubated overnight at 37 °C. DNA extraction was performed from 400 μl of broth with MagNa Pure Compact Nucleic Acid isolation Kit I according to manufacturers' protocol version 12 for DNA extraction from bacteria. An Illumina HiSeq platform with a 2 × 100 paired end run was used for WGS. The paired reads and merging contigs were assembled by Geneious version 8.1.5. and the MIRA plugin 1.0.1 (Kearse et al., 2012). Only sequences with a coverage of >70 were proceeded. The core genome ANI was calculated using the Gegenees software version 2.2.1 with blast plugin. A threshold of 20% was chosen to make sure that only the core genomes were compared (Ågren et al., 2012). The result file was transferred as *.next file to SplitsTree4 version 4.13.1 (Huson and Bryant, 2006) to visualize the results as a phylogenetic tree. The workflow of WGS-ANI analysis is shown in Fig. 1.

2.3. Pulse-field gel electrophoresis (PFGE)

PFGE of all VRE-isolates was carried out at the clinical microbiological laboratory of the Public Health Agency of Sweden. The abbreviations of the PFGE clusters consisted of 5 to 6 sections in the locally developed nomenclature at the agency: SE = Sweden, Efm = *Enterococcus faecium*, the resistance gene *vanA* or *vanB*, the year when the cluster was detected for the first time, and a serial number for instance SE-EfmA-1410. A lowercase letter after the serial number (SE-EfmA-1410a) indicated that the band pattern was >90% but <97% similar to the base cluster (SE-EfmA-1410).

2.4. Multilocus sequence typing (MLST)

MLST was performed *in silico* using the WGS data. The online platform tool MLST 1.8 (Larsen et al., 2012) was used to determine the MLST types.

3. Results

3.1. Epidemiological investigation

The IPC team detected epidemiological links between 37 patients (38 isolates) in three separate outbreaks between 2013 and 2015 involving seven different wards. During 2013–2014 a total of 29 patients with *vanB* had epidemiological links and had been transferred between five wards. The 29 patients were suspected to have acquired VRE in one medical ward (15 patients), one surgical ward (three patients), one geriatric ward (six patients), one elderly home (two patients), and a second elderly home (three patients) (Table 1). During 2014 a total of five patients with *vanA* were suspected to have acquired VRE in a cardiologic ward and during 2015 a total of six patients with *vanB* were suspected to have acquired VRE in a medical ward.

3.2. Microbiological investigation

>14,000 screening samples were analysed at the clinical microbiological laboratory of UUH between 2013 and 2015 of which 10% resulted in positive gene detection for *vanA* or *vanB* gene. Since other species than *E. faecium* and *E. faecalis* may contain *vanB* genes, both the selective cultivation and phenotypic verification of the isolates had to be positive to define a sample as VRE positive. Out of all *vanA* or *vanB* positive samples 5% were characterised phenotypically as enterococci by Maldi-Tof. Out of 14,000 screening samples 49 isolates of *E. faecium* with *vanB* gene and 11 isolates of *E. faecium* with *vanA* gene were detected.

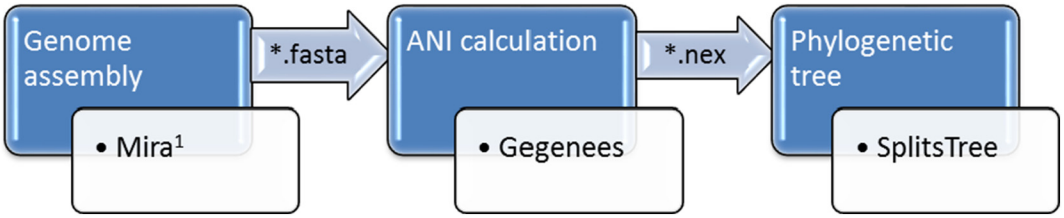

Fig. 1. Workflow for WGS-ANI determination. [1]Velvet was also tested with the same result (results not shown).

Table 1
Summary of all sampling data (BioProject PRJNA301929); ANI- standard sequences are marked bold.

Outbreak and ward	Gene	Lab.no.	NCBI accession	Sex	Age	Sampling date	ANI cluster	Result PFGE	MLST
Outbreak 1 medical ward 2014	*van B*	E13931	LNLB00000000	M	71	n.a.	5	SE-EfmB-1308	ST-192
		S-1001508	LNLC00000000	M	65	2010		SE-EfmB-0701	ST-192
		VRE-1300911	**LNLJ00000000**	F	52	2013–12-27		SE-EfmB-1308	ST-192
		VRE-1300937	LNLK00000000	M	95	2013–12-29		SE-EfmB unique	ST-192
		VRE-1400136	LNLL00000000	M	70	2014–01-09		SE-EfmB-1308d	ST-192
		VRE-1400373	LNLO00000000	F	19	2014–01-27		SE-EfmB-1308f	unknown ST
		VRE-1400408	LNLP00000000	F	49	2014–01-27		SE-EfmB-1308b	ST-192
		VRE-1400413	LNLQ00000000	M	67	2014–01-27		SE-EfmB unique	ST-192
		VRE-1401098	LNLR00000000	M	80	2014–02-24		SE-EfmB-1308	ST-192
		VRE-1401338	LNLT00000000	F	51	2014–03-10		SE-EfmB-1308f	ST-192
		VRE-1402215	LNLZ00000000	M	30	2014–04-14		SE-EfmB-1308b	ST-192
		VRE-1403299	LNMK00000000	F	64	2014–05-12		SE-EfmB-1308	ST-192
		VRE-1404669	LNMP00000000	F	86	2014–06-16		SE-EfmB-1308	ST-192
		VRE-1502382	LNOT00000000	F	54	2015–06-07		SE-EfmB-1308	ST-192
		VRE-1402006	LNLY00000000	F	79	2014–04-07	3	SE-EfmB unique	ST-78
Outbreak 1 surgical ward	*van B*	**VRE-1400294**	**LNLM00000000**	F	87	2014–01-20	2	SE-EfmB-1402	ST-117
		VRE-1400325	LNLN00000000	M	75	2014–01-23		SE-EfmB-1402	
		VRE-1401318	LNLS00000000	M	89	2014–03-07		SE-EfmB-1402b	
Outbreak 1 geriatric ward	*van B*	VRE-1401988	LNLX00000000	F	94	2014–04-08	4	SE-EfmB-1308d	ST-317
		VRE-1402253	LNMB00000000	F	87	2014–04-15			
		VRE-1402258	LNMC00000000	F	87	2014–04-14			
		VRE-1402259	LNMD00000000	F	73	2014–04-15			
		VRE-1402435	LNME00000000	F	83	2014–04-22			
		VRE-1402673	**LNMI00000000**	M	92	2014–04-28			
Outbreak 1 elderly home 1	*van B*	VRE-1401379	LNLU00000000	F	95	2014–03-11	5	SE-EfmB-1308	ST-192
		VRE-1401859	LNLV00000000	F	91	2014–03-26		SE-EfmB-1308b	
		VRE-1401878	LNLW00000000	F	90	2014–04-01		SE-EfmB-1308	
Outbreak 1 elderly home 2	*van B*	**VRE-1402513**	**LNMF00000000**	M	79	2014–04-23	6	SE-EfmB unique	unknown ST
		VRE-1403540	LNMM00000000	M	85	2014–05-16	5	SE-EfmB-1308d	ST-192
Outbreak 2 cardiologic ward	*van A*	VRE-1407687	LNMS00000000	M	85	2014–11-21	1	SE-EfmA-1410	ST-80
		VRE-1407988	LNMT00000000	M	79	2014–12-01			
		VRE-1408197	LNMV00000000	M	79	2014–12-10			
		VRE-1408429	**LNMW00000000**	F	96	2014–12-16			
		VRE-1408535	LNMX00000000	M	71	2014–12-19			
Outbreak 3 medical ward 2015	*van B*	**VRE-1502856**	**LNOU00000000**	F	81	2015–07-11	3	SE-EfmB-1509	ST-117
		VRE-1502913	LNOV00000000	F	90	2015–07-14			
		VRE-1503262	LNOX00000000	F	84	2015–07-28			
		VRE-1503268	LNOY00000000	F	95	2015–07-28			
		VRE-1503642	LNOZ00000000	F	75	2015–08-05			
		VRE-1503646	LNPA00000000	M	77	2015–08-05			
Cases with epidemiological links to Gavle outbreak	*van A*	U-1313438	LNLE00000000	M	49	2013–11-17	No cluster assignment	SE-EfmA unique	ST-203
		VRE-1300899	LNLH00000000	M	82	2013–12-22			ST-18
		VRE-1300900	LNLI00000000	n.a.	82	2013–12-23			ST-80
		VRE-1402237	LNMA00000000	M	75	2014–04-15			ST-787
		VRE-1402563	LNMG00000000	M	81	2014–04-23			ST-721
		VRE-1406033	LNMQ00000000	F	52	2014–06-06			ST-203
	van B	87,056,200	LNDL00000000	M	64	n.a.	2	SE-EfmB-1402a	ST-117
		VRE-1502939	LNOW00000000	M	73	2015–07-16	6	SE-EfmB unique	unknown ST
		197,806,558	LNLA00000000	M	41	n.a.	5	SE-EfmB-1308	ST-192
		S-1402282	LNLD00000000	F	61	2014–05-07		SE-EfmB-1308a	
		VRE-1300518	LNLF00000000	F	82	2013–10-24		SE-EfmB-1308b	
		VRE-1300578	LNLG00000000	F	56	2013–11-03		SE-EfmB-1308b	
		VRE-1402601	LNMH00000000	M	79	2014–04-26		SE-EfmB-1308b	
		VRE-1402991	LNMJ00000000	M	66	2014–05-06		SE-EfmB-1308	
		VRE-1403355	LNML00000000	F	68	2014–05-13		SE-EfmB-1308b	
		VRE-1404029	LNMN00000000	F	69	2014–05-28		SE-EfmB-1308e	
		VRE-1404192	LNMO00000000	F	59	2014–06-02		SE-EfmB-1308	
		VRE-1406092	LNMR00000000	F	61	2014–09-12		SE-EfmB-1308	
		VRE-1408033	LNMU00000000	M	78	2014–12-02		SE-EfmB unique	
		VRE-1504220	LNPB00000000	M	43	2015–08-22		SE-EfmB unique	

n.a. no data available; unknown ST means an allelic profile without assigned ST.

3.3. Multilocus sequence typing (MLST)

The MLST analysis revealed that 58 out of 60 *E. faecium* VRE isolates could be assigned to one out of nine known MLST types and only two isolates remained with an allelic profile without assigned ST = unknown ST (Table 1). The predominant ST in this study was ST-192 (n = 29) which corresponds to 50% of all assigned isolates. Less frequent types were ST-18, ST-78, ST-787 (n = 1), ST-721, ST-203 (n = 2) and ST-203 (n = 2).

3.4. Pulsed-field gel electrophoresis (PFGE)

Altogether 46 *E. faecium* VRE isolates could be assigned to one of the 12 PFGE groups (Table 1). The remaining isolates had a unique PFGE pattern with no PFGE cluster assignment.

Out of 38 VRE isolates that were suspected to belong to the three outbreaks, eight separate clusters were identified by PFGE-analysis. The PFGE clustering of the first outbreak, involving five wards between 2013 and 2014 was not congruent with the epidemiological

investigation. The 29 isolates were characterised as SE-EfmB-1308 (n = 8), SE-EfmB-1308b (n = 3), SE-EfmB-1308d (n = 8), SE-EfmB-1308f (n = 2), SE-EfmB-1402 (n = 2), SE-EfmB-1402b (n = 1), SE-EfmB-701 (1) and four isolates remained "unique" The PFGE pattern designations are added to the ANI tree in Fig. 2. All five isolates that belonged to the second outbreak in 2014 in the cardiologic ward were characterised as SE-EFmA-1410. All six isolates that belonged to the third outbreak in 2015 in the medical ward were characterised as SE-EfmB-1509.

3.5. Whole genome sequencing (WGS)

Altogether 60 *E. faecium* VRE-strains of genotype *vanA* and *vanB* were sequenced with an Illumina platform and the genomes were assembled on scaffold level. Gap closing was not performed since it is not applicable for clinical diagnostic approaches and has no influence on the current analysis.(Greub et al., 2009) The final assemblies of all *E. faecium* VRE strains are available from the NCBI database BioProject number PRJNA301929 (Table 1).

The core genome based WGS ANI analysis divided 53 out of 60 isolates into six clusters; ANI1 (n = 5), ANI2 (n = 4), ANI3 (n = 6), ANI4 (n = 6), ANI5 (n = 30), and ANI6 (n = 2), see Figs. 2 and 3. The remaining seven isolates did not belong to a cluster according to the WGS analysis and were defined as "unique".

The WGS clustering had high accordance to the epidemiological investigation. Out of 38 VRE *E. faecium* isolates involved in the three outbreaks with clear epidemiologic links, the WGS analysis identified five clusters and two isolates were defined as "unique". The 29 isolates that belonged to the first outbreak between 2013 and 2014 in five wards belonged to ANI5 (n = 19 from the medical ward and n = 1 from an elderly home), ANI4 (n = 6 in the geriatric ward), ANI2 (n = 3 in the surgical ward). The two isolates from patients in the second elderly home did not cluster together in WGS at all; one isolate clustered with ANI5 and one with ANI6. All five isolates that belonged to the second outbreak in 2014 in the cardiologic ward belonged to ANI1. All six isolates that belonged to the third outbreak in 2015 in the medical ward belonged to ANI3.

The interval of the intraspecific divergence of the whole genomes of the 60 examined *E. faecium* VRE isolates was 0.1% (VRE-1406092 to 197,806,558) to 4.4% (VRE-1406033 to 197,806,558).

3.6. Comparison of PFGE, MLST and WGS-ANI

For the ANI based table of distances (ToD) the whole genomes of all 60 strains were compared to all others, which resulted in 3600 ANI results. The ANIs of all isolates that either belong to the PFGE cluster SE-EfmB-1402, SE-EfmB-1402a, SE-EfmB-1402b, SE-EfmB-1410, or SE-EfmB-1509 are outlined in Table 2 as example for the outbreak-specific cut-off point determination. The lowest ANI of each cluster was calculated and compared to the highest ANI values between those PFGE clusters. The cluster SE-EfmB-1509 had a lowest internal ANI of 98.7%. This means that the highest genetic divergence between 2 strains within this cluster is 1.3% while the lowest genetic divergence between two isolates from several clusters SE-EfmB-1509 and SE-EfmB-1402a is 0.85%. The results of all ANI comparisons based on ANI cluster, PFGE cluster, and on the MLST type is summarized in Table 3. Due to the high discriminatory power of WGS-ANI it was possible to divide ST-117 in two clusters while ST-80, ST-317 and ST-192 are representing an ANI cluster each.

The cut-off point interval for the genetic divergence of the whole genomes within all examined clusters (Δ ANI_{max}) was in the interval of 0.5% to 1.5% for PFGE, 1.2% to 2.1% in MLST and 1.04 to 1.44% in WGS-ANI analysis. ANI cluster 6 was ignored, since it consisted of only two isolates and therefore was not representative. However, this cluster showed the advantage of the automatic clustering using WGS-ANI. While the ANI of those isolates confirmed that the strains belong to

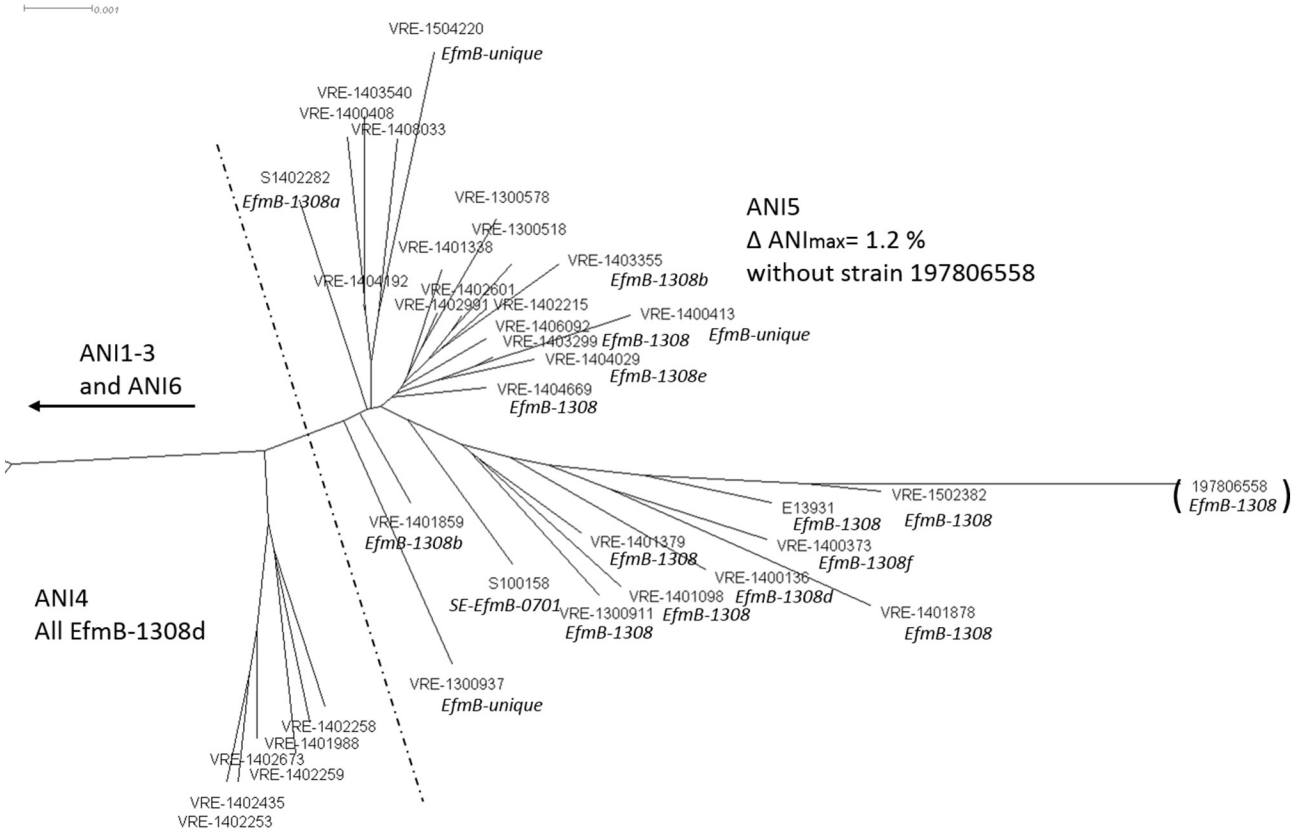

Fig. 2. WGS-ANI cluster 4 and 5; selected PFGE cluster assignment results were added in italics. Strain 197806558 was not directly connected to the recent outbreaks.

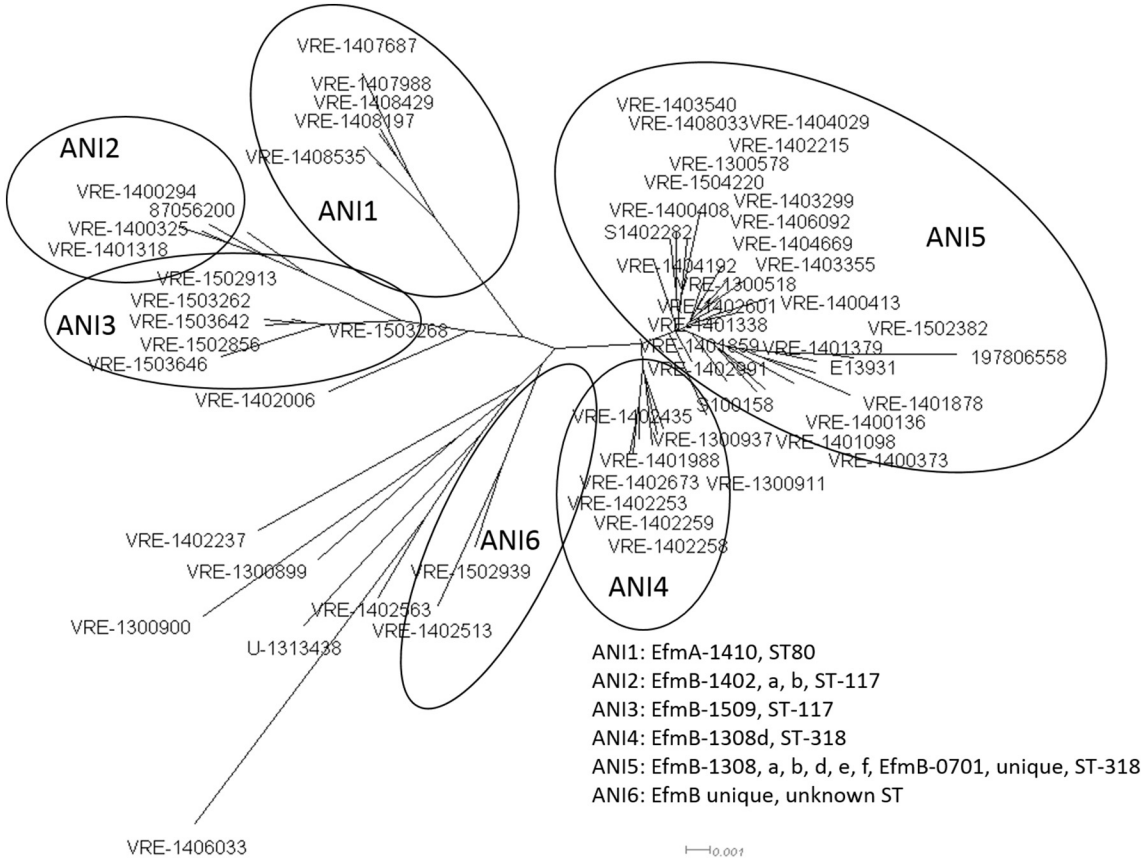

Fig. 3. Complete phylogenetic tree of all investigated strains and classification of all outbreak related VRE strains in 6 WGS-ANI clusters.

the same cluster (ANI = 99.35%), the strains remained as unique patterns in PFGE and as unknown ST in MLST.

4. Discussion

Nosocomial outbreaks of VRE are an ever present challenge to the health-care system. Infection prevention and control departments must collaborate closely with the clinical microbiological laboratory and appropriate molecular typing methods must be used in order to confirm or reject cases with epidemiological links. Molecular methods for genotyping with high discriminatory ability to compare isolates are crucial. In this study we analysed 60 *E. faecium* VRE-isolates from three minor nosocomial outbreaks between 2013 and 2015 in the county of Uppsala, Sweden with PGFE, WGS and MLST. The aim of this study was to examine the applicability of a new WGS-ANI workflow compared to the established methods PFGE and MLST for genotyping isolates in an outbreak situation.

Discrepancies between the PFGE results and the epidemiological investigation were observed during the first *vanB* outbreak between 2013 and 2014 involving 29 patients in five wards but confirmed the epidemiological investigation in the other two outbreaks, in the cardiological ward during 2014 and in the medical ward during 2015. In total PFGE identified six clusters among all 38 isolates. Those clusters defined by PFGE were not in accordance to the epidemiological definition of a

Table 2
Whole genome ANI table of the PFGE clusters SE-EfmB-1402, SE-EfmB-1402a, SE-EfmB-1402b, SE-EfmB-1410, or SE-EfmB-1509.

MLST	PFGE cluster	Sample number	1	2	3	4	5	6	7	8	9	10	11	12	13	14
ST-114	SE-EfmB-1402	1: VRE-1400294	100	99.1	99.3	99.6	97.8	97.7	97.1	97.8	97.4	98.7	98.8	98.1	98.1	98.3
		2: VRE-1400325	99.3	100	99.3	99.7	97.7	97.7	97.1	97.8	97.4	98.7	98.7	98	98	98.3
	SE-EfmB-1402a	3: 87056200	99.1	99	100	99.5	97.9	98	97.6	98.2	97.9	99	99.1	98.5	98.6	98.9
	SE-EfmB-1402b	4: VRE-1401318	99.1	99	99.3	100	97.6	97.6	97	97.7	97.3	98.5	98.6	98	97.9	98.2
ST-80	SE-EfmA-1410	5: VRE-1407687	97.6	97.4	97.9	97.9	100	99.3	98.6	99.3	98.9	97.5	97.6	97	96.8	97.1
		6: VRE-1407988	97.8	97.6	98.3	98.1	99.6	100	99.2	99.8	99.5	98	98.1	97.5	97.5	97.7
		7: VRE-1408197	97.8	97.7	98.3	98.1	99.8	99.8	100	99.8	99.8	98	98.1	97.9	97.9	98
		8: VRE-1408429	97.8	97.6	98.3	98.1	99.5	99.7	99.2	100	99.5	98	98.1	97.5	97.4	97.8
		9: VRE-1408535	97.9	97.8	98.4	98.2	99.7	99.8	99.6	99.8	100	98.1	98.2	97.8	97.8	98
ST-117	SE-EfmB-1509	10: VRE-1502856	98.8	98.6	99.2	99	97.7	97.9	97.4	98	97.8	100	99.7	99	99.1	99.3
		11: VRE-1502913	98.8	98.6	99.2	99	97.8	98	97.5	98.1	97.8	99.7	100	99.1	99.1	99.4
		12: VRE-1503268	98.9	98.8	99.3	99.2	98.1	98.1	98	98.2	98.1	99.7	99.8	100	99.5	99.6
		13: VRE-1503642	98.8	98.7	99.3	99.1	97.8	97.9	97.9	98	97.9	99.6	99.7	99.4	100	99.5
		14: VRE-1503646	98.5	98.3	99	98.8	97.5	97.7	97.2	97.8	97.5	99.3	99.4	98.7	98.7	100

Table 3
Whole genome ANI comparisons based on ANI cluster, PFGE cluster, and to the MLST type; bold: upper threshold of the method.

		n ANI pairs	ANI Average within the cluster	ANI standard deviation	ANI min	ANI max	Δ ANI max	PFGE cluster	MLST type
ANI cluster	ANI 1	20	99.50%	0.35%	98.56%	99.84%	**1.44%**	SE-EfmA-1410	ST-80
	ANI 2	12	99.27%	0.23%	98.96%	99.66%	1.04%	SE-EfmB-1402, a, b	ST-117
	ANI 3	30	99.42%	0.35%	98.69%	99.92%	1.31%	SE-EfmB-1509	ST-117
	ANI 4	30	99.40%	0.30%	98.80%	99.80%	1.20%	SE-EfmB-1308d	ST-317
	ANI 5	870	99.40%	0.30%	98.80%	99.90%	1.20%	SE-EfmB-1308, a,b,d,e,f	ST-192
	ANI 6	2	99.35%	–	99.35%	99.35%	0.65%	vanB unique	unknown ST
PFGE cluster	SE-EfmB-1308	132	99.40%	0.40%	98.50%	99.90%	1.50%	–	ST-192
	SE-EfmB-1308b	42	99.50%	0.20%	99.00%	99.80%	1.00%	–	ST-192
	SE-EfmB-1308d	56	99.20%	0.30%	98.50%	99.80%	**1.50%**	–	ST-192, ST-317
	SE-EfmB-1308f	2	99.50%	–	99.50%	99.50%	0.50%	–	ST-192, unknown ST
	SE-EfmB-1402	2	99.20%	–	99.20%	99.20%	0.80%	–	ST-117
	SE-EfmB-1509	20	99.40%	0.30%	98.70%	99.80%	1.30%	–	ST-117
	SE-EfmA-1410	20	99.50%	0.30%	98.60%	99.80%	1.40%	–	ST-80
MLST type	ST-117	90	99.02%	0.49%	97.90%	99.90%	2.10%	SE-EfmB-1402, a, SE-EfmB-1509	–
	ST-80	30	98.90%	0.97%	96.70%	99.80%	2.10%	SE-EfmA-1410, vanA unique	–
	ST-203	2	97.00%	–	97.00%	97.00%	**3.00%**	vanA unique	–
	ST-317	30	99.40%	0.30%	98.80%	99.80%	1.20%	SE-EfmB-1308d	–
	ST-192	812	99.33%	0.38%	98.00%	99.90%	2.00%	SE-EfmB-1308, a,b,d,e,f, unique, SE-EfmB-0701	–

cluster in the first outbreak. For example, in the medical ward where 12 patients which were suspected to have acquired VRE over a short period of time. PFGE characterised the isolates as belonging to four different clusters with no clear explanation; SE-EfmB-1308, SE-EfmB-1308b, SE-EfmB-1308d, SE-EfmB-1308f, and "unique".

For the two other outbreaks involving 5 VRE *vanA* in the cardiologic ward 2014 and 6 VRE *vanB* in the medical ward 2015 PFGE confirmed the epidemiological investigation. Thus, a test for the applicability of whole genome sequencing approaches for epidemiological source tracing was carried out. The WGS analysis of the 38 isolates had high accordance with the epidemiological investigation. Based on the WGS results, the IPC team had useful information about which patients belonged to the chain of transmission. The IPC team could then proceed with the proper interventions to stop further transmission in the wards.

Most of the published studies about WGS applications for epidemiological investigation are based on analysis of single nucleotide polymorphisms/variants (SNP, SNV) (Kinnevey et al., 2016; Sherry et al., 2013) or on the comparison of sequence fragments that were derived from WGS data. Salipante et al. (2015) described a WGS based method for comparison of whole VRE genomes. In those studies, the data analysis was based on a manual one by one sequence blast which limits the usability of the method due to the number of comparable strains. To overcome the one by one comparison of sequences, the Gegenees software was used in our study to calculate the table of distances (Ågren et al., 2012). The software enables adding new sequences to an existing database. To assign isolates to outbreaks, a cut off point for outbreak strain delineation as it exists for common species delineation (ANI = 94% (Richter and Rossello-Mora, 2009)) was calculated. Out of this study, a threshold of 98.5% is recommended for *E. faecium* VRE strain outbreak cluster delineation in WGS-ANI analysis. An easy to interpret alternative to a defined cut off point is an ANI cluster visualization using a phylogenetic tree (Figs. 1 and 2).

PFGE is still the gold standard for molecular epidemiological investigation of nosocomial outbreaks of VRE. PFGE is cost effective and the direct clustering of the band pattern in PFGE is a user friendly advantage. However, PFGE analysis has several limitations. Even if an image of the PFGE gel is analysed by a software, subjective ocular examination of bands in the gel which might occur shifted or weak is often necessary and there exist a risk of human mistakes. Increasing the resolution of PFGE is challenging. When PFGE cluster SE-EFm-1308 was divided into several subtypes to get a better resolution, the results of the epidemiological investigation and the subtyping of SE-EFm-1308 had low accordance.

MLST types can be determined from WGS data. Thus, the results of WGS-ANI can be compared with existing MLST data. MLST has a higher cut-off point (2.1%) regarding outbreak strain delineation than PFGE (1.5%) and WGS-ANI (1.44%) and the strain assignment is therefore very accurate in MLST. Despite the effect that ST-1170 was divided in two ANI clusters the MLST results are conform to the ANI results. The lower discriminatory power of MLST becomes a disadvantage if a higher resolution is needed as is the case in nosocomial outbreaks. While the threshold of resolution in MLST and PFGE is determined by the applied enzymes, the threshold in WGS-ANI is adaptable depending on the examined time interval. The compared strains of the presented study are isolated in a time interval of 5 years (S-1001508 = 2010 to VRE-1503646 = 2015). If the ANI of strains that are isolated within a long time interval should be calculated, the evolutionary clock speed of *E. faecium* may cause a wrong result. Two isolates VRE-1300911 and VRE-1502382, that were sampled in a time interval of 18 months from the same patient showed an ANI of 99.3%. Both isolates probably represented the same strain as they were isolated from the same patient, they belonged to the same PFGE cluster SE-EfmB-1308 and to the same sequence type (ST-192).

4.1. Accuracy of WGS ANI

The maximal resolution of the WGS-ANI method is determined by the number of nucleotides in the whole genome and is theoretically about $1{:}10^6$ and $1{:}3 \times 10^6$ for bacteria with genome sizes of 1 Mb and 3 Mb, respectively. Salipante et al. (2015) determined a technical error of the Illumina sequencing method including all steps from library preparation to bioinformatics of 0.467 ± 0.333 ($n = 19$).

4.2. Data sharing

The ANI calculation of a new database takes only a few minutes or at maximum a few hours if many whole genomes are included. Once a genome is available from the public genome databases the ANI database calculation can be done by each laboratory. The ANI results of different outbreak investigations could become comparable using one cluster standard for each ANI cluster or at least one "standard" genome. Out of our study we recommend using the sequences: LNMW00000000 (ANI1), LNLM00000000 (ANI2), LNOU00000000 (ANI3), LNMI00000000 (ANI4), LNMF00000000 (ANI5), and LNLJ00000000 (ANI6) as standards for the ANI clusters (Table 1).

5. Conclusion

WGS-ANI is an easy to use method for genotyping VRE which is applicable in daily diagnostics and which allows to share the data worldwide. It is a more user friendly method compared to MLST and PFGE. WGS-ANI has an increased discriminatory power as MLST and PFGE and a better epidemiological concordance as PFGE. NGS based approaches minimize the risk of human mistakes which still is a problem in PFGE. WGS data can be used for further studies beyond outbreak investigations, such as detection of resistance- or virulence- genes.

As a result of this study we recommend using the described WGS-ANI workflow (Fig. 1) instead of PFGE for epidemiological outbreak investigations. The recommended cut-off point for ANI based VRE outbreak-cluster delineation is 98.5%. The method is not limited to the analysis of VRE and can be used for epidemiological issues for other species as well.

Acknowledgements

We are grateful to Cecilia Svensson for the help with WGS. Furthermore we thank Karolina Gullsby and Anki Pahv for providing VRE strains from Gaevle and Eskilstuna.

References

Ågren, J., Sundström, A., Håfström, T., Segerman, B., 2012. Gegenees: fragmented alignment of multiple genomes for determining phylogenomic distances and genetic signatures unique for specified target groups. PLoS One 7, e39107.

de Been, M., Pinholt, M., Top, J., Bletz, S., Mellmann, A., van Schaik, W., Brouwer, E., Rogers, M., Kraat, Y., Bonten, M., Corander, J., Westh, H., Harmsen, D., Willems, R.J.L., 2015. Core genome multilocus sequence typing scheme for high-resolution typing of *Enterococcus faecium*. J. Clin. Microbiol. 53, 3788–3797.

van Belkum, A., 1994. DNA fingerprinting of medically important microorganisms by use of PCR. Clin. Microbiol. Rev. 7, 174–184.

Cookson, B.D., Robinson, D.A., Monk, A.B., Murchan, S., Deplano, A., de Ryck, R., Struelens, M.J., Scheel, C., Fussing, V., Salmenlinna, S., Vuopio-Varkila, J., Cuny, C., Witte, W., Tassios, P.T., Legakis, N.J., van Leeuwen, W., van Belkum, A., Vindel, A., Garaizar, J., Haeggman, S., Olsson-Liljequist, B., Ransjo, U., Muller-Premru, M., Hryniewicz, W., Rossney, A., O'Connell, B., Short, B.D., Thomas, J., O'Hanlon, S., Enright, M.C., 2007. Evaluation of molecular typing methods in characterizing a European collection of epidemic methicillin-resistant *Staphylococcus aureus* strains: the harmony collection. J. Clin. Microbiol. 45, 1830–1837.

Greub, G., Kebbi-Beghdadi, C., Bertelli, C., Collyn, F., Riederer, B.M., Yersin, C., Croxatto, A., Raoult, D., 2009. High throughput sequencing and proteomics to identify immunogenic proteins of a new pathogen: the dirty genome approach. PLoS One 4, e8423.

Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254–267.

Kao, R.R., Haydon, D.T., Lycett, S.J., Murcia, P.R., 2014. Supersize me: how whole-genome sequencing and big data are transforming epidemiology. Trends Microbiol. 22, 282–291.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., Drummond, A., 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.

Kinnevey, P.M., Shore, A.C., Mac Aogáin, M., Creamer, E., Brennan, G.I., Humphreys, H., Rogers, T.R., O'Connell, B., Coleman, D.C., 2016. Enhanced tracking of nosocomial transmission of endemic sequence type 22 methicillin-resistant *Staphylococcus aureus* type IV isolates among patients and environmental sites by use of whole-genome sequencing. J. Clin. Microbiol. 54, 445–448.

Konstantinidis, K.T., Tiedje, J.M., 2005. Genomic insights that advance the species definition for prokaryotes. Proc. Natl. Acad. Sci. U. S. A. 102, 2567–2572.

Larsen, M.V., Cosentino, S., Rasmussen, S., Friis, C., Hasman, H., Marvig, R.L., Jelsbak, L., Sicheritz-Ponten, T., Ussery, D.W., Aarestrup, F.M., Lund, O., 2012. Multilocus sequence typing of total-genome-sequenced bacteria. J. Clin. Microbiol. 50, 1355–1361.

Maiden, M.C., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D.A., Feavers, I.M., Achtman, M., Spratt, B.G., 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. U. S. A. 95, 3140–3145.

Richter, M., Rossello-Mora, R., 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc. Natl. Acad. Sci. U. S. A. 106, 19126–19131.

Ruiz-Garbajosa, P., Bonten, M.J.M., Robinson, D.A., Top, J., Nallapareddy, S.R., Torres, C., Coque, T.M., Cantón, R., Baquero, F., Murray, B.E., del Campo, R., Willems, R.J.L., 2006. Multilocus sequence typing scheme for *Enterococcus faecalis* reveals hospital-adapted genetic complexes in a background of high rates of recombination. J. Clin. Microbiol. 44, 2220–2228.

Salipante, S.J., SenGupta, D.J., Cummings, L.A., Land, T.A., Hoogestraat, D.R., Cookson, B.T., 2015. Application of whole-genome sequencing for bacterial strain typing in molecular epidemiology. J. Clin. Microbiol. 53, 1072–1079.

Sherry, N.L., Porter, J.L., Seemann, T., Watkins, A., Stinear, T.P., Howden, B.P., 2013. Outbreak investigation using high-throughput genome sequencing within a diagnostic microbiology laboratory. J. Clin. Microbiol. 51, 1396–1401.

Tenover, F.C., Arbeit, R.D., Goering, R.V., Mickelsen, P.A., Murray, B.E., Persing, D.H., Swaminathan, B., 1995. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. J. Clin. Microbiol. 33, 2233–2239.

Valdezate, S., Labayru, C., Navarro, A., Mantecon, M.A., Ortega, M., Coque, T.M., Garcia, M., Saez-Nieto, J.A., 2009. Large clonal outbreak of multidrug-resistant CC17 ST17 *Enterococcus faecium* containing Tn5382 in a Spanish hospital. J. Antimicrob. Chemother. 63, 17–20.

Werner, G., 2013. Molecular typing of enterococci/VRE. J. Bacteriol. Parasitol. S5-001 (doi 10, 2155-9597).

Werner, G., Klare, I., Fleige, C., Geringer, U., Witte, W., Just, H.M., Ziegler, R., 2012. Vancomycin-resistant vanB-type *Enterococcus faecium* isolates expressing varying levels of vancomycin resistance and being highly prevalent among neonatal patients in a single ICU. Antimicrob. Resist. Infect. Control 1, 21.