

RESEARCH

Open Access



Long-read sequencing-based in silico phage typing of vancomycin-resistant *Enterococcus faecium*

Paola Lisotto¹, Erwin C. Raangs¹, Natacha Couto^{1,2}, Sigrid Rosema¹, Mariëtte Lokate¹, Xuewei Zhou¹, Alexander W. Friedrich¹, John W. A. Rossen^{1,3,4}, Hermie J. M. Harmsen¹, Erik Bathoorn^{1*} and Monika A. Chlebowicz-Fliss¹

Abstract

Background: Vancomycin-resistant enterococci (VRE) are successful nosocomial pathogens able to cause hospital outbreaks. In the Netherlands, core-genome MLST (cgMLST) based on short-read sequencing is often used for molecular typing. Long-read sequencing is more rapid and provides useful information about the genome's structural composition but lacks the precision required for SNP-based typing and cgMLST. Here we compared prophages among 50 complete *E. faecium* genomes belonging to different lineages to explore whether a phage signature would be usable for typing and identifying an outbreak caused by VRE. As a proof of principle, we investigated if long-read sequencing data would allow for identifying phage signatures and thereby outbreak-related isolates.

Results: Analysis of complete genome sequences of publicly available isolates showed variation in phage content among different lineages defined by MLST. We identified phage present in multiple STs as well as phages uniquely detected within a single lineage. Next, in silico phage typing was applied to twelve MinION sequenced isolates belonging to two different genetic backgrounds, namely ST117/CT24 and ST80/CT16. Genomic comparisons of the long-read-based assemblies allowed us to correctly identify isolates of the same complex type based on global genome architecture and specific phage signature similarity.

Conclusions: For rapid identification of related VRE isolates, phage content analysis in long-read sequencing data is possible. This allows software development for real-time typing analysis of long-read sequencing data, which will generate results within several hours. Future studies are required to assess the discriminatory power of this method in the investigation of ongoing outbreaks over a longer time period.

Keywords: MinION, Outbreak, VRE, Typing, Prophage, Molecular epidemiology

Background

The prevalence of vancomycin-resistant *Enterococcus faecium* (VRE_{fm}) carriage among hospitalized patients has increased dramatically over the past 15 years. In

many parts of Europe, this has resulted in a rapid increase in infections with vancomycin-resistant enterococci (VRE) [1]. This is mainly due to the nosocomial spread of VRE_{fm} lineages that are highly adapted to survive in the hospital environment [2]. This is worrisome for several reasons. VRE may cause severe infections in immune-compromised patients, such as patients with hematologic disorders and solid organ transplant

* Correspondence: d.bathoorn@umcg.nl

¹Department of Medical Microbiology and Infection Prevention, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands
Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

patients [3]. Since there is a minimal choice of alternative antibiotic treatment options, VRE has been categorized as a high priority pathogen in the WHO list of antibiotic-resistant pathogens for which the development of new antibiotics is required [4].

To control the spread of VRE f_m , it is recommended to treat patients carrying VRE f_m with barrier precautions, trace contacts, and type the isolates [5]. Patients who carry VRE f_m can develop infections with this pathogen later [6]. Therefore, it makes sense to prevent the transmission of VRE by screening for carriage [7]. Molecular typing of isolates shows whether they are closely related and thereby indicates transmission or not. In case genetically related isolates are identified, contact investigations are initiated. The typing methods used must have a rapid turnaround time for a fast response in outbreak management. Next to this, high discriminatory power is required for VRE f_m typing. Due to the rapid emergence of nosocomial VRE f_m lineages, the chromosomal diversity within these lineages is relatively low. This makes it difficult to rule-in isolates that are involved in an outbreak. Even SNP-based analysis of next generation sequencing data may lack power to prove transmission. Nonetheless, core-genome MLST (cgMLST) based on short-read sequencing is recommended for molecular typing of VRE f_m , since it is a validated method with higher discriminatory power compared to other typing methods as PFGE and MLST [8, 9]. The cgMLST typing results of isolates can be generated in 24–72 h, counting from the moment of DNA extraction [10].

Long-read sequencing is faster and provides useful information about the structural composition of the genome. Results can be generated within minutes [11]. It can be used to detect VRE through metagenomics [11] and for MLST and resistance gene detection in whole-genome sequences of isolates [12]. A drawback of MinION long-read sequencing is the higher error rate when compared to MiSeq short-read sequencing. Since around 8% of bases are miscalled, nanopore sequencing thus far lacks the precision required for SNP-based typing methods [13]. Information about the structural genome composition, however, can be useful for typing as well. Mobile genetic elements may be reliably identified in nanopore sequencing data. This is relevant for VRE outbreaks since recent studies have shown that the type of transposons and their insertion sites in the genome can be specific for an outbreak clone [14–16].

The characterization of bacteriophage content may also be useful for typing. The use of phages in the typing of bacteria had been used in the past. Back in the 1930s, *Salmonella* species were typed by studying which phages were able to infect the specific strains and to which phages the isolate is resistant [17]. Later on, this phage

typing method was also developed for many other pathogens such as *Staphylococcus aureus* [18] and *Enterococcus* species [19] and used for outbreak investigations. It was noticed that the resistance pattern to infection by bacteriophages was related to the presence of certain prophages [20]. The presence of specific prophage gives immunity to the infections caused by a similar group of bacteriophages. Thus, the conventional phage typing was an indirect detection method for strain-specific prophages. The recent developments in long-read sequencing allow for direct and more robust sequence reconstruction of prophages and their integration sites in bacterial genomes [21]. Thus, direct sequencing information about phage content is more informative than that derived by conventional phage typing. The gain and loss of phages are critical drivers in the differentiation of subtypes with strain-specific properties, even in lineages that are highly similar in their core genome [22]. This leads to our hypothesis that long-read-based characterization of phage content may have high discriminatory power in identifying outbreak strains.

Here, we compared prophages among complete genome sequences of genetically diverse *E. faecium* isolates to explore whether a signature of phage content would be usable for typing of this pathogen. Next, as a proof of principle, we investigated if long-read sequencing data allows for identifying such signatures in clinical outbreak situations by using data from well described outbreak strains [15, 16].

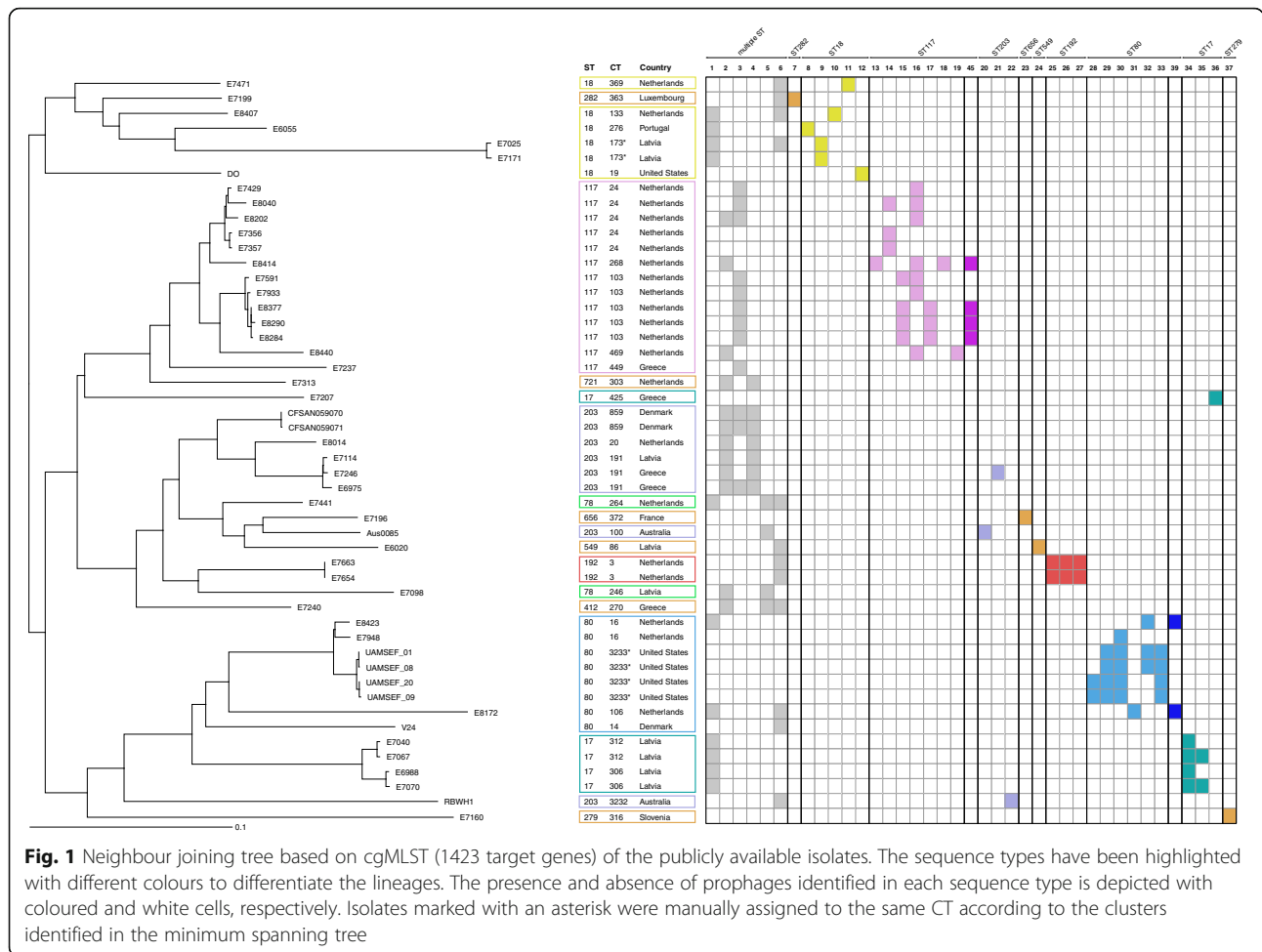
Results

Prophage profiles of the 50 publicly available *E. faecium* genomes

In the first part of this study, we investigated the presence of prophages among complete genomes of *E. faecium* isolates from different geographical and genetic backgrounds. In this analysis, we aim to identify prophage patterns that are associated with specific lineages.

In Fig. 1, we show the phylogenetic relatedness of 50 *E. faecium* genomes from hospitalised patients, belonging to different sequence types of clonal complex 17 (CC17). For additional information regarding the phylogenetic relationship of these strains, we refer to Additional file 2 where a minimum spanning tree indicates the allelic differences between the isolates.

A variable number of prophages were identified in individual isolates, with a maximum of five prophage regions per genome. When compared to the NCBI viral database, no significant alignments were found that exceeded 10% coverage. An overview of the phages detected in the investigated isolates and their sequence annotations, are presented in Additional files 1 and 3. Six phages (Phage 1 to 6) were detected in multiple lineages while the majority of the



phages (Phage 7 to 37) were uniquely present within a single sequence type. The phylogenetic analysis (Additional file 4) revealed that these phages cluster in two major groups, and one phage clustered separately. The first group comprises five phages, which are closest related to the *Siphoviridae* Bacillus SPbeta and Staphylococcus phage SPbeta-like. The second group includes the majority of the phages ($n = 44$) which are closest related to a clade of unclassified *Siphoviridae*, mainly detected in *Listeria* spp. Only Phage 17 is recognized as distinct phage, that is closest relative to the Enterococcus phage IME_EF of the *Siphoviridae* family.

From this analysis, we learned that it is not feasible to type VRE isolates using a single prophage as a reference for typing of outbreak events, given the high diversity in prophage content among different lineages. Instead, it seemed useful to characterise the total prophage content of the index isolate of any given outbreak and compare sequences of other isolates to such unique prophage signature to investigate outbreak events.

Twelve *E. faecium* isolates for the use of in silico phage typing

In the second part of the study, we applied in silico prophage typing to 12 nosocomial isolates from the UMCG. These isolates belonged to two different sequence and complex types, namely ST117, CT24 and ST80, CT16. Epidemiological data related to these isolates is summarised in Table 1.

To assess these isolates' phylogenetic relatedness, cgMLST was performed on the MiSeq data (Fig. 2). Among isolates of CT16 we detected 11 allele differences, when considering samples collected in the same outbreak, only 3 allele differences were identified. In isolates from CT24, we detected a maximum of 8 allele differences.

Among investigated UMCG isolates, a total of eight prophages (phage 38–45) were identified based on long-read data. These were used to create a concatenated reference sequence for the BLAST analysis presented in Fig. 3 and summarised in Table 1. To verify that the profile observed was unique for UMCG isolates, this

Table 1 Information on the UMCG isolates used in this study

ID	V1	V2	BL4	BL5	BL6	B2	B11	A7	BL1	BL2	BL3	S1
ST	80	80	80	80	80	117	117	117	117	117	117	117
CT	16	16	16	16	16	24	24	24	24	24	24	24
Vancomycin Resistance	VRE	VRE	VSE	VSE	VSE	VRE	VRE	VRE	VSE	VSE	VSE	VSE
Year of collection	2019	2019	2016	2015	2014	2017	2017	2014	2017	2015	2014	2016
Phage 38	1	1	0	0	0	0	0	0	0	0	0	0
Phage 39	1	1	1	1	0	0	0	0	0	0	0	0
Phage 40	1	1	1	1	1	0	0	0	0	0	0	0
Phage 41	0	0	0	0	1	0	0	0	0	0	0	0
Phage 42	0	0	0	0	0	1	1	1	1	1	1	1
Phage 43	0	0	0	0	0	1	1	1	1	1	1	1
Phage 44	0	0	0	0	0	1	0	0	1	1	1	1
Phage 45	0	0	0	0	0	0	0	1	0	0	0	0
Phage 40.1	1	1	1	1	1	1	1	1	1	1	1	1
Phage 42.1	0	0	0	0	0	1	1	1	1	1	1	1
Phage 44.1	0	0	0	0	0	1	1	0	0	0	0	0

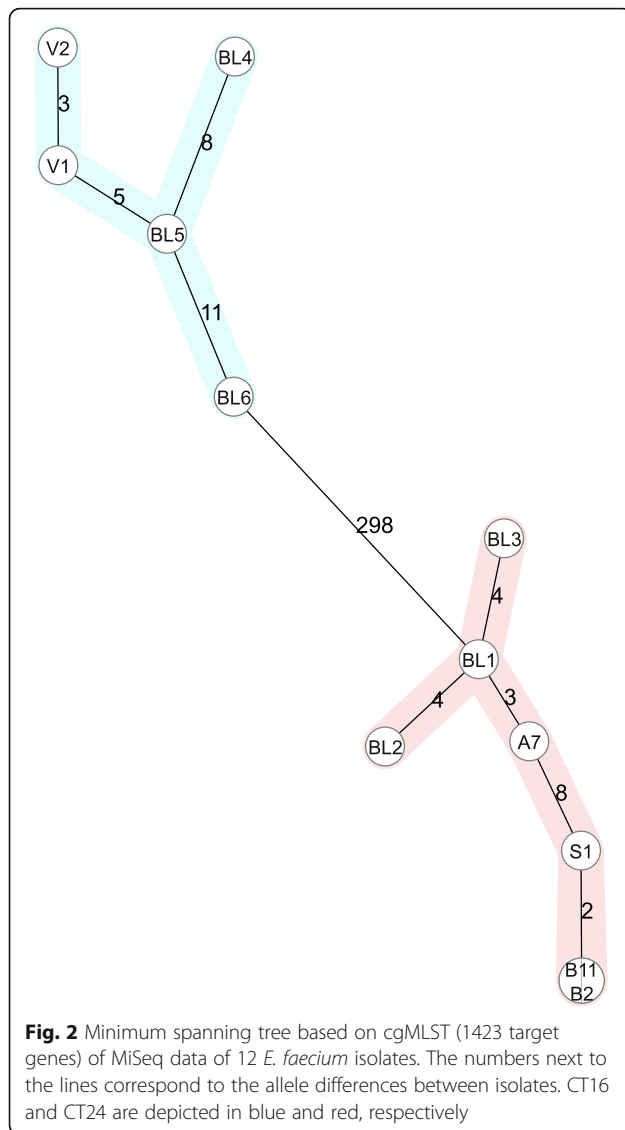
Abbreviations: ID identification, ST sequence type, CT complex type.

reference phage sequence was compared to the publicly available isolates. Phage 38 and 41 were uniquely present in UMCG isolates V1, V2 and BL6, respectively. Phage 39 and 45 were also present in genomes of publicly available isolates belonging to the same sequence type (Fig. 1). Phages 40, 42, 43 and 44 presented a high similarity with phages 17, 22, 3 and 24, respectively.

Next, as a proof-of-principle, we tested the use of long-read sequencing based in silico phage identification for the typing of *E. faecium* in an outbreak investigation. We exclusively used isolate B11 which belonged to the index patient of an outbreak event and identified the phages that characterise this isolate (Phage 43, 40.1, 42.1, 44.1). We then used these phages to generate a concatenated reference sequence and compared it to isolate B2, collected during the same outbreak episode (Fig. 4A). Despite some minor gaps, the sequence coverage of the reference was 98%, indicating a high similarity between isolates. We investigated in which functional module of the predicted prophage sequence the main differences occurred among the closely related isolates of ST117/CT24 presented in Fig. 4B. This revealed that all the isolate which did not belong to the 2017 outbreak differed in the *integrase* gene, and in an adjacent cluster of hypothetical protein genes. Isolate B2 slightly differed from the reference in the endonuclease and recombinase genes. In this set of isolates, the median reference coverage was found to be lowered to 96%. Interestingly, one isolates (A7) collected during a different outbreak showed a low similarity (78%) due to the lack of phage 44.1. This classification is in line with the cgMLST

analysis presented in Fig. 3 and with the SNP-based analysis presented in Additional file 5. To further evaluate our method, we expanded the analysis by including 57 isolates collected from VREfm outbreaks that occurred in 2014 and 2017. The data presented in Additional file 6 confirmed that isolates collected during 2017 outbreak belonging to ST117/CT24 lineage showed the highest similarity to the reference isolate B11 (97–100% coverage). Instead, the isolates collected in 2014 outbreak and classified as ST117/CT24 did not show a concordant profile, having a median reference coverage of 79%. Comparable or even lower similarity was observed when isolates obtained from 2014 outbreak but belonging to different lineages, e.i. ST80/CT104 (50%) and ST117/CT103 (79%), were included. When representative isolates belonging to different sequence types and collected outside our hospital were used for the comparison (Fig. 4C), the median coverage was all < 70%, underscoring their unrelatedness. The analysis of the complete list of isolates is presented in Additional files 7 and 8.

Finally, we evaluated the proposed phage typing method using short-reads data in isolates typed as ST80/CT104, collected during an outbreak event in 2014. In this lineage, two phages were identified and compared to the same 57 investigated isolates where the median coverage for this reference was 95%. Using this approach all outbreak isolates belonging to ST80/CT104 were correctly identified with a coverage of 100% (Additional file 9). Nearly all isolates of other tested outbreaks showed coverage below 97%. The main differences observed



between the reference and other isolates were linked to some elements of the phage structure related to the replication, tail, packing and lysogeny modules (Additional file 10). However, a cluster of isolates belonging to a different complex type (ST117/CT103) showed phage sequences with high percentages coverage (96–100%), unexpectedly.

According to our results, a threshold of 97% coverage could be considered, allowing to rule-out clonal transmission. The phage typing method cannot be used for a definitive inclusion of isolates into an outbreak cluster, since we found a cluster of false positive matches from a different lineage. The phage typing method provides complementary results to typing by cgMLST, and can therefore be used as tool to discriminate between isolates that belong to the same complex type, but are not epidemiologically related.

Discussion

Enterococcus faecium has been widely recognised as a potential threat for hospitalised patients, especially immune-compromised, due to the multidrug resistance profile characterizing this opportunistic pathogen [4]. Early detection of VRE is a crucial element in preventing hospital outbreaks [2]. In this scenario, the use of next-generation sequencing (NGS) has been proven successful and allowed to track the transmission route with variable turnaround time depending on the technology used [10]. In this study, we applied in silico phage typing using MinION long-read sequencing as an alternative approach for the early detection and screening of VRE_{fm} outbreaks.

In our analysis, a variable number of prophages were found to be integrated in the genome of different *E. faecium* lineages. This is in line with previous observations indicating that bacterial genomes contain up to 20% of prophage genes [23]. Van Schaik et al. estimated that for *E. faecium* strains, between 2 and 5% of the total number of CDS could belong to predicted phages [24]. To the best of our knowledge, no other similar research was performed, indicating that this is the first study to identify prophage signatures as a typing method for *E. faecium*. Our analysis of publicly available isolates revealed the presence of prophages shared between isolates belonging to different lineages and unique profiles present within the same lineage.

The application of solely long-read sequencing for the assembly of complete genomes has proved successful for different microorganisms [21, 25], and recently the MLST genotyping of *E. faecium* using ONT technology was suggested in hospital infection control [12]. Our study confirmed the possibility of *E. faecium* full genome reconstruction using only long-read data which supports the possibility to identify a prophage signature, possibly consisting of multiple phages for a higher discriminatory power, that could be applied during an outbreak event as a typing method to identify if the isolates are related. Our analysis on long-read sequencing data indicates that, when comparing isolates sharing the same genetic background as predicted by cgMLST, a high degree of similarity can be found. Contrary, if different lineages are compared, a clear distinction of prophage profile can be observed. This indicates that, in a situation where clonal spread is present, a shared prophage reference could identify related isolates and distinguish them from unrelated ones. We observed that using only two phages as reference can be a limitation of this method, not allowing to rule out clonal transmission especially when comparing isolates collected during a long period of time. Recent horizontal transfer of bacteriophages also may result in the detection of identical phage sequences in different lineages. None of the cgMLST, SNP-based

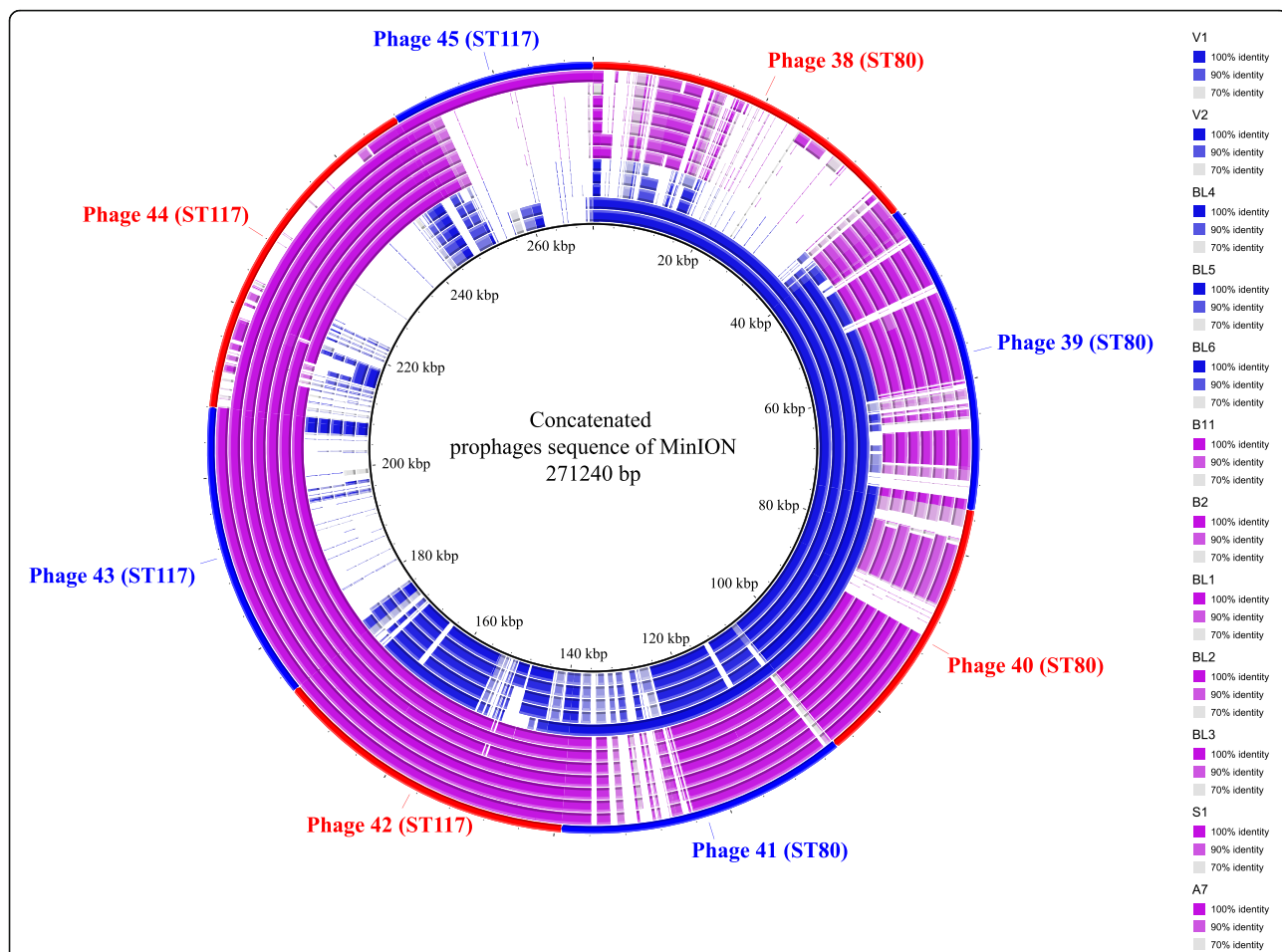


Fig. 3 Concatenated prophage sequences identified in UMCG isolates. The concatenated reference sequence is indicated as the red-blue outer circle, the inner circles represent the UMCG isolates. CT16 and CT24 isolates are depicted in dark-blue and purple, respectively. The isolates in the legend should be read starting from top to down and from inner circle to outer circle

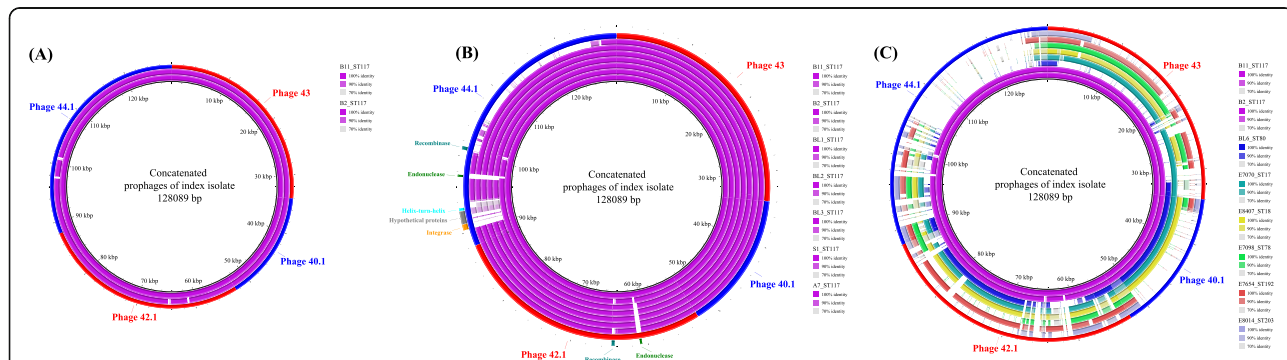


Fig. 4 (A-C): Clinical application of prophage typing during an outbreak investigation. The concatenated prophage reference sequence is indicated as red-blue outer circle. In panel A, the first inner circle from the centre represents the sequence of an outbreak isolate (index patient) belonging to ST117/CT24. The second circle from the centre represents the sequence of an additional isolate which belongs to the same outbreak event and complex type. In panel B, next to the isolates indicated in panel A, isolates belonging to the same complex type, but not epidemiologically related to the same outbreak event are presented. In the last outer ring, the genes responsible for the differences between the isolates has been annotated. In panel C, sequences of additional isolates belonging to different sequence types than the outbreak event are depicted in different colours. The isolates in the legend should be read starting from top to down and from inner circle to outer circle

and the in silico phage typing methods have the discriminatory power to prove clonal transmission conclusively, however, the results of all these methods provide complementary information, and would need to be interpreted together with epidemiological data.

Interestingly, we detected genetic diversity in the integrase gene in closely related isolates belonging to CT24. In literature [26, 27], this gene was described as a potential target for typing and as an indicator of genomic diversity, making it a useful target to distinguish between closely related strains within hyperepidemic lineages.

The application of third-generation sequencing, such as ONT technology, to clinical microbiology and infection prevention is of great potential [28]. The fast turnaround time and the possibility to analyse preliminary data in real-time make this technology a promising approach for outbreak investigations where a quick reaction is essential [25]. Computed prediction of prophage regions, and direct detection of these regions on the sequence reads would strongly shorten the turnaround time. In that case, we estimate that results could be generated in less than 2 h. We underline that applying this technique would be more suitable in short-term outbreak investigation considering the variability in phage content that can occur over time. An example of such scenario was isolate A7 which, due to the lack of phage 44.1, had a lower similarity to the reference compared to other isolates of the same CT. Moreover, we acknowledge that the use of only long-read sequencing data should be further developed and tested in future studies with a larger sample size to validate our results.

Conclusions

In conclusion, we show a proof of principle for in silico prophage typing. Based on this method, we were able to group CT24 *E. faecium* isolates and distinguish them from different complex types. Using this approach we correctly identified isolates belonging to the same outbreak event, which is not always possible based on cgMLST. Long-read sequence-based phage typing is promising as a rapid typing method with high discriminatory power, and it seems worthwhile to assess this in future validation studies.

Methods

E. faecium isolates

For this study, 50 hybrid assembled *E. faecium* genomes were downloaded from NCBI or the publicly available Gitlab repository (https://gitlab.com/sirarredondo/efaecium_population/tree/master/Files/Unicycler_assemblies). These isolates were collected from hospitalised patients, and their metadata is summarised in Additional file 1.

Additionally, twelve *E. faecium* isolates were used, collected at the University Medical Center Groningen (UMCG) between 2014 and 2019. Five of these isolates (A7, B2, B11, V1, V2) were collected during three different outbreak events in 2014, 2017, and 2019. The other isolates were either recovered from independent bloodstream infections (BL1-BL6) or surveillance procedures (S1). These *E. faecium* isolates were initially sequenced with Illumina MiSeq and retrospectively assessed by Oxford Nanopore Technologies (ONT) sequencing (MinION).

Finally, VREfm isolates collected in two outbreak events which occurred in 2014 and 2017 and previously described by Zhou et al. and Lisotto et al. have been included in this study to further validate the method proposed [15, 16].

DNA preparation and sequencing

Twelve bacterial isolates were grown overnight on blood agar plates at 37 °C. Bacterial DNA isolation was performed using the DNeasy UltraClean Microbial Kit (MO BIO Laboratories, Carlsbad, CA) following the manufacturer's protocol. DNA concentration and purity were measured by the Qubit dsDNA HS and BR assay kit (Life Technologies, Carlsbad, CA, USA). Initially, these isolates were sequenced by short-read sequencing with Illumina MiSeq (Illumina, San Diego, CA, USA). DNA libraries were prepared with the Nextera XT v2 kit (Illumina). These isolates were also sequenced with the MinION (ONT, Oxford, UK). Libraries were prepared using the Rapid Barcoding Sequencing kit, SQK-RBK004 (ONT). The library was loaded on a FLO-MIN106D flow cell R9.4.1 (ONT) and sequenced on a MinION device for 48 h. Base-calling was performed using Guppy v2.1.3 (ONT).

Data analysis

De novo assemblies of short-read sequencing were performed in CLC Genomics Workbench v12 (QIAGEN, Hilden, Germany). The adapters of the long-reads were trimmed and demultiplexed by using Porechop v0.2.4 (<https://github.com/rrwick/Porechop>). Long-read assemblies were performed by Flye v2.5 [29]. The cgMLST typing was performed using Ridom SeqSphere+ software v6.0.3 (Ridom GmbH, Münster, Germany). Snippy v4.3.0 (<https://github.com/tseemann/snippy>) was used to identify core genome concatenated single nucleotide polymorphisms (SNPs) in the short-read sequenced isolates using *E. faecium* E1 strain (NZ_CP018065.1) as reference.

Phage identification

The presence of prophages was investigated using the online tool PHASTER [30]. The predicted prophage regions were manually extracted from the genomic

sequences and were investigated in detail using Artemis [31]. We investigated the genome of *E. faecium*, considering the direction of the open reading frames genes in the predicted prophage regions as suggested before [23, 32]. In the outbreak investigation, a concatenated reference sequence was created containing the prophage genomes identified in *E. faecium* isolate of the index patient. Subsequently, a BLAST analysis was performed to compare the reference against the long-read sequencing data of other isolates. In this analysis we accepted some degree of sequence discrepancy when comparing assemblies generated from long-read sequencing data. The results were visualised using BRIG [33]. Moreover, ViPTree [34] was used to classify the bacteriophages in a genome-wide approach.

Abbreviations

VRE: Vancomycin-resistant enterococci; VRE_{fm}: Vancomycin-resistant *Enterococcus faecium*; cgMLST: Core-genome Multi Locus Sequence Type; UMCG: University Medical Center Groningen; ONT: Oxford Nanopore Technologies; E1: *E. faecium* strain (NZ_CP018065.1); CC17: Clonal complex 17; CT: Complex type; ST: Sequence type; NGS: Next-generation sequencing

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-021-08080-5>.

Additional file 1. Accession number and information on publicly available isolates.

Additional file 2. Minimum spanning tree based on cgMLST (1423 target genes) of the publicly available isolates. Numbers within the circle indicate the isolates. The numbers next to the lines correspond to allele differences between the isolates. The different colours highlight groups of closely related isolates.

Additional file 3. Sequence annotation of the prophages regions identified in the isolates investigated in the study.

Additional file 4. Dendrograms revealing genomic similarity in the publicly available isolates.

Additional file 5 SNP-based analysis of the short-read data from UMCG isolates. *E. faecium* E1 strain (NZ_CP018065.1) was used as reference.

Additional file 6. Analysis of the UMCG outbreak isolates against concatenated reference sequence of index isolate (B11). The concatenated prophage reference sequence is indicated as red-blue outer circle. The isolates in the legend should be read from top to down and from inner circle to outer circle.

Additional file 7. Analysis of the publicly available isolates against the concatenated reference sequence of index isolate (B11). The isolates in the legend should be read from top to down and from inner circle to outer circle.

Additional file 8. Percentage of similarity between the publicly available genomes and the concatenated reference sequence of index isolate (B11).

Additional file 9. Percentage of similarity between the outbreak isolates and the reference sequence of ST80/CT104 lineage.

Additional file 10. Analysis of the outbreak isolates against the reference sequence of ST80/CT106 lineage. The isolates in the legend should be read from top to down and from inner circle to outer circle.

Acknowledgements

Not applicable.

Authors' contributions

ER performed the experiments. ML was responsible for the clinical case. PL, NC, SR, MC analysed the data. PL prepared the figures and wrote the manuscript. JR, HH, EB, MC commented on the interpretation of the results. NC, XZ, AF, JR, HH revised the manuscript. EB, MC conceived the study and participated in its coordination, helped to draft the manuscript, and supervised the study throughout. All authors contributed to and approved the final manuscript.

Funding

This study was supported by the Marie Skłodowska-Curie Actions (Grant Agreement Number: 713660 - PRONKJEWAIL - H2020-MSCA-COFUND-2015).

Availability of data and materials

Illumina MiSeq and Oxford Nanopore Technologies MinION reads of the 12 *E. faecium* isolates collected at UMCG and used in this study have been deposited in the European Nucleotide Archive (ENA) public project PRJEB41626 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB41626>). Isolate A7 can be retrieved under the BioProject no PRJEB25590; BioSample SAMEA104703752 (<https://www.ebi.ac.uk/ena/browser/view/SAMEA104703752>).

Declarations

Ethics approval, accordance and consent to participate

The samples used for the present analyses were collected during routine diagnostics and infection prevention and control investigations. All procedures were carried out according to guidelines and regulations of UMCG concerning the use of patient materials for the validation of clinical methods, which are in compliance with the guidelines of the Federation of Dutch Medical Scientific Societies (FDMSS). Every patient entering the UMCG is informed that samples taken may be used for research and publication purposes, unless they indicate that they do not agree to it. This protocol has been approved by the Medical Ethical Committee of the UMCG and registered under the project number 201900812. Written informed consent was obtained from all individuals or their guardians prior to study participation. All samples were used after performing and completing a conventional microbiological diagnosis and were coded to protect patients' confidentiality. All experiments were performed in accordance with the guidelines of the Declaration of Helsinki and the institutional regulations.

Consent for publication

Not applicable.

Competing interests

JR is currently employed by IDbyDNA Inc. The other authors declare that they have no competing interests.

Author details

¹Department of Medical Microbiology and Infection Prevention, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. ²The Milner Centre for Evolution, Department of Biology and Biochemistry, University of Bath, Bath, UK. ³Department of Pathology, University of Utah School of Medicine, Salt Lake City, UT, USA. ⁴IDbyDNA Inc., Salt Lake City, UT, USA.

Received: 4 March 2021 Accepted: 7 October 2021

Published online: 23 October 2021

References

1. ECDC. Surveillance atlas of infectious diseases (2018). <https://atlas.ecdc.europa.eu/public/index.aspx> [Accessed 16 Aug 2018].
2. Zhou X, Willems RJL, Friedrich AW, Rossen JWA, Bathoorn E. Enterococcus faecium: from microbiological insights to practical recommendations for infection control and diagnostics. Antimicrob Resist Infect Control. 2020;9(1): 130. <https://doi.org/10.1186/s13756-020-00770-1>.
3. Nellore A, Huprikar S, AST ID Community of Practice. Vancomycin-resistant Enterococcus in solid organ transplant recipients: Guidelines from the American Society of Transplantation Infectious Diseases Community of Practice. Clin Transpl. 2019;33(9):e13549. <https://doi.org/10.1111/ctr.13549>.

4. Tacconelli E, Magrini N. Global Priority List of Antibiotic-Resistant Bacteria to Guide Research, Discovery, and Development of New Antibiotics. In: Tacconelli E, Carrara E, Savoldi A, Kattula D, Burkert F, editors. Organización Mundial de La Salud: World Health Organization; 2017. <http://www.cdc.gov/drugresistance/threat-report-2013/>.
5. Siegel JD, Rhinehart E, Jackson M, Chiarello L, and the Healthcare Infection Control Practices Advisory Committee, 2007 Guideline for Isolation Precautions: Preventing Transmission of Infectious Agents in Healthcare Settings <https://www.cdc.gov/infectioncontrol/guidelines/isolation/index.html>
6. Rubin IMC, Pedersen MS, Mollerup S, Kaya H, Petersen AM, Westh H, et al. Association between vancomycin-resistant *Enterococcus faecium* colonization and subsequent infection: a retrospective WGS study. *J Antimicrob Chemother.* 2020;75(7):1712–5. <https://doi.org/10.1093/jac/dkaa074>.
7. Johnstone J, Shing E, Saedi A, Adomako K, Li Y, Brown KA, et al. Discontinuing contact precautions for vancomycin-resistant *Enterococcus* (VRE) is associated with rising VRE bloodstream infection rates in Ontario hospitals, 2009–2018: a quasi-experimental study. *Clin Infect Dis.* 2020;71(7): 1756–9. <https://doi.org/10.1093/cid/ciaa009>.
8. de Been M, Pinholt M, Top J, Bletz S, Mellmann A, van Schaik W, et al. Core Genome Multilocus Sequence Typing Scheme for High-Resolution Typing of *Enterococcus faecium*. *J Clin Microbiol.* 2015;53(12):3788–97. <https://doi.org/10.1128/JCM.01946-15>.
9. Lytsy B, Engstrand L, Gustafsson Å, Kaden R. 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. *Infect Genet Evol.* 2017; 54:74–80. <https://doi.org/10.1016/j.meegid.2017.06.010>.
10. Deurenberg RH, Bathoorn E, Chlebowicz MA, Couto N, Ferdous M, García-Cobos S, et al. Application of next generation sequencing in clinical microbiology and infection prevention. *J Biotechnol.* 2017;243:16–24. <https://doi.org/10.1016/j.jbiotec.2016.12.022>.
11. Yee R, Breitwieser FP, Hao S, Opene BNA, Workman RE, Tamma PD, et al. Metagenomic next-generation sequencing of rectal swabs for the surveillance of antimicrobial-resistant organisms on the Illumina MiSeq and Oxford MinION platforms. *Eur J Clin Microbiol Infect Dis.* 2021;40(1):95–102. <https://doi.org/10.1007/s10096-020-03996-4>.
12. Tarumoto N, Sakai J, Sujino K, Yamaguchi T, Ohta M, Yamagishi J, et al. Use of the Oxford Nanopore MinION sequencer for MLST genotyping of vancomycin-resistant enterococci. *J Hosp Infect.* 2017;96(3):296–8. <https://doi.org/10.1016/j.jhin.2017.02.020>.
13. Jain M, Olsen HE, Paten B, Akeson M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol.* 2016;17(1):239. <https://doi.org/10.1186/s13059-016-1103-0>.
14. Bender JK, Kalmbach A, Fleige C, Klare I, Fuchs S, Werner G. Population structure and acquisition of the vanB resistance determinant in German clinical isolates of *Enterococcus faecium* ST192. *Sci Rep.* 2016;6:21847. <https://doi.org/10.1038/srep21847>.
15. Zhou X, Chlebowicz MA, Bathoorn E, Rosema S, Couto N, Lokate M, et al. Elucidating vancomycin-resistant *Enterococcus faecium* outbreaks: the role of clonal spread and movement of mobile genetic elements. *J Antimicrob Chemother.* 2018;73(12):3259–67. <https://doi.org/10.1093/jac/dky349>.
16. Lisotto P, Couto N, Rosema S, Lokate M, Zhou X, Bathoorn E, et al. Molecular characterisation of vancomycin-resistant *Enterococcus faecium* isolates belonging to the lineage ST117/CT24 causing hospital outbreaks. *Front Microbiol.* 2021. <https://doi.org/10.3389/fmicb.2021.728356>.
17. Craigie J. A bacteriophage for the vi-form of *Salmonella typhi*. *J Bacteriol.* 1936;31:56.
18. Anderson ES, Williams RE. Bacteriophage typing of enteric pathogens and staphylococci and its use in epidemiology. *J Clin Pathol.* 1956;9(2):94–127. <https://doi.org/10.1136/jcp.9.2.94>.
19. Caprioli T, Zaccour F, Kasatiya S. Phage typing scheme for group D streptococci isolated from human urogenital tract. *J Clin Microbiol.* 1975;2: 311–7.
20. Brock TD. Host range of certain virulent and temperate bacteriophages attacking. *J Bacteriol.* 1964;88:165–71. <https://doi.org/10.1128/JB.88.1.165-171.1964>.
21. González-Escalona N, Allard MA, Brown EW, Sharma S, Hoffmann M. Nanopore sequencing for fast determination of plasmids, phages, virulence markers, and antimicrobial resistance genes in Shiga toxin-producing *Escherichia coli*. *PLoS One.* 2019;14(7):e0220494. <https://doi.org/10.1371/journal.pone.0220494>.
22. van Schaik W, Willems RJ. Genome-based insights into the evolution of enterococci. *Clin Microbiol Infect.* 2010;16(6):527–32. <https://doi.org/10.1111/j.1469-0691.2010.03201.x>.
23. Casjens S. Prophages and bacterial genomics: what have we learned so far? *Mol Microbiol.* 2003;49(2):277–300. <https://doi.org/10.1046/j.1365-2958.2003.03580.x>.
24. van Schaik W, Top J, Riley DR, Boekhorst J, Vrijenhoek JE, Schapendonk CM, et al. Pyrosequencing-based comparative genome analysis of the nosocomial pathogen *Enterococcus faecium* and identification of a large transferable pathogenicity island. *BMC Genomics.* 2010;11:239. <https://doi.org/10.1186/1471-2164-11-239>.
25. Quick J, Ashton P, Calus S, Chatt C, Gossain S, Hawker J, et al. Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of *Salmonella*. *Genome Biol.* 2015;16:114. <https://doi.org/10.1186/s13059-015-0677-2>.
26. Crestani C, Forde TL, Zadoks RN. Development and application of a prophage integrase typing scheme for group B *Streptococcus*. *Front Microbiol.* 2020;11:1993. <https://doi.org/10.3389/fmicb.2020.01993>.
27. Colavecchio A, D'Souza Y, Tompkins E, Jeukens J, Freschi L, Emond-Rheault JG, et al. Prophage integrase typing is a useful indicator of genomic diversity in *Salmonella enterica*. *Front Microbiol.* 2017;8:1283. <https://doi.org/10.3389/fmicb.2017.01283>.
28. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* 2020; 21(1):30. <https://doi.org/10.1186/s13059-020-1935-5>.
29. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol.* 2019;37(5):540–6. <https://doi.org/10.1038/s41587-019-0072-8>.
30. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 2016; 44(W1):W16–21. <https://doi.org/10.1093/nar/gkw387>.
31. Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. Artemis: an integrated platform for visualisation and analysis of high-throughput sequence-based experimental data. *Bioinformatics.* 2012;28(4):464–9. <https://doi.org/10.1093/bioinformatics/btr703>.
32. Yasmin A, Kenny JG, Shankar J, Darby AC, Hall N, Edwards C, et al. Comparative genomics and transduction potential of *Enterococcus faecalis* temperate bacteriophages. *J Bacteriol.* 2010;192(4):1122–30. <https://doi.org/10.1128/JB.01293-09>.
33. Ali Khan NF, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics.* 2011;12:402. <https://doi.org/10.1186/1471-2164-12-402>.
34. Nishimura Y, Yoshida T, Kuronishi M, Uehara H, Ogata H, Goto S. ViPTree: the viral proteomic tree server. *Bioinformatics.* 2017;33(15):2379–80. <https://doi.org/10.1093/bioinformatics/btx157>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

