A Core Genome Multilocus Sequence Typing Scheme for *Enterococcus faecalis*

Bernd Neumann,[a] Karola Prior,[b] Jennifer K. Bender,[a] Dag Harmsen,[b] Ingo Klare,[a] Stephan Fuchs,[a] Astrid Bethe,[c] Daniela Zühlke,[d] André Göhler,[e] Stefan Schwarz,[c] Kirsten Schaffer,[f] Katharina Riedel,[d] Lothar H. Wieler,[g] Guido Werner[a]

[a]Robert Koch Institute, Division of Nosocomial Pathogens and Antibiotic Resistances, Wernigerode, Germany
[b]Department of Periodontology and Restorative Dentistry, University Hospital Muenster, Muenster, Germany
[c]Institute of Microbiology and Epizootics, Centre for Infection Medicine, Freie Universität Berlin, Berlin, Germany
[d]Department for Microbial Physiology and Molecular Biology, Institute of Microbiology, University of Greifswald, Greifswald, Germany
[e]Friedrich Loeffler Institute of Medical Microbiology, University Medicine Greifswald, Greifswald, Germany
[f]Department of Microbiology, St. Vincent's University Hospital and School of Medicine and Medical Science, University College Dublin, Dublin, Ireland
[g]Robert Koch Institute, Berlin, Germany

ABSTRACT Among enterococci, *Enterococcus faecalis* occurs ubiquitously, with the highest incidence of human and animal infections. The high genetic plasticity of *E. faecalis* complicates both molecular investigations and phylogenetic analyses. Whole-genome sequencing (WGS) enables unraveling of epidemiological linkages and putative transmission events between humans, animals, and food. Core genome multilocus sequence typing (cgMLST) aims to combine the discriminatory power of classical multilocus sequence typing (MLST) with the extensive genetic data obtained by WGS. By sequencing a representative collection of 146 *E. faecalis* strains isolated from hospital outbreaks, food, animals, and colonization of healthy human individuals, we established a novel cgMLST scheme with 1,972 gene targets within the Ridom SeqSphere$^+$ software. To test the *E. faecalis* cgMLST scheme and assess the typing performance, different collections comprising environmental and bacteremia isolates, as well as all publicly available genome sequences from the NCBI and SRA databases, were analyzed. In more than 98.6% of the tested genomes, >95% good cgMLST target genes were detected (mean, 99.2% target genes). Our genotyping results not only corroborate the known epidemiological background of the isolates but exceed previous typing resolution. In conclusion, we have created a powerful typing scheme, hence providing an international standardized nomenclature that is suitable for surveillance approaches in various sectors, linking public health, veterinary public health, and food safety in a true One Health fashion.

KEYWORDS *Enterococcus faecalis*, core genome MLST, molecular surveillance, molecular typing, whole-genome sequencing

Enterococcus faecalis is a ubiquitous environmental and opportunistic Gram-positive bacterium that colonizes the gastrointestinal tract of humans and various animals (1). *E. faecalis* is also used in commercial probiotic products and is an important ingredient in food production, such as in fermented sausages or cheese from raw milk (2–5). Among members of the genus *Enterococcus*, this species shows the highest incidence for human and animal infections and, in addition, is discussed as a zoonotic pathogen (5–7). *E. faecalis* has emerged as a major nosocomial pathogen worldwide, causing bloodstream infections mainly in immunocompromised humans. Furthermore, it is responsible for endocarditis and urinary tract infections acquired on an outpatient basis, as well as for mastitis in dairy cattle (8–10). Due to their ability to survive even harsh environmental conditions, enterococci pose an immense hygienic challenge to

clinical settings (11–13). Most importantly, treatments of enterococcal infections are limited to only a few therapeutic agents, since enterococci exhibit multiple intrinsic resistances and are known to easily acquire additional resistance determinants (14, 15). Hence, multidrug-resistant enterococci, especially vancomycin-resistant enterococci (VRE), are a burden for hospitals worldwide (16, 17).

To validate epidemiological linkages and disclose possible transmission events among humans, animals, food, and the environment, discriminatory high-resolution typing is of the utmost importance. The high genetic plasticity of *E. faecalis* is the result of wide exchanges of genetic material, which in turn complicate molecular investigations (18). In contrast to the second most important *Enterococcus* species in hospital settings, *Enterococcus faecium*, for which hospital-adapted lineages have already been identified, most *E. faecalis* genotypes do not show extended host or context specificity (19). For instance, Buhnik-Rosenblau et al. demonstrated, on the basis of multilocus sequence typing (MLST), that no specific genetic groups could be assigned to a colonization- or infection-associated origin (5). Nevertheless, other studies have described at least one hospital-associated *E. faecalis* lineage, to which sequence type 6 (ST6) belongs (20–23).

Classical MLST relies on PCR amplification and the sequencing of seven housekeeping genes that are distributed over the bacterial chromosome (23, 24). Classical MLST provides an international and expandable typing nomenclature, but the method provides only a moderate typing resolution. Typing by macrorestriction analysis via pulsed-field gel electrophoresis (PFGE) is limited in its discriminatory power, and it sometimes displays discrepancies with whole-genome sequencing (WGS)-based investigations (25, 26). Since next-generation sequencing (NGS) enables high-throughput analyses of entire bacterial genomes at an affordable cost, it has quickly become indispensable for performing population and outbreak analyses. However, appropriate bioinformatics tools are necessary as a prerequisite for handling and interpreting sequence data (25, 27, 28). Core genome MLST (cgMLST) aims to combine the discriminatory power of classical MLST with the extensive genetic data derived from WGS (29, 30). Exploiting hundreds of gene targets of the entire bacterial genome, thereby providing maximum resolution for multiple research and surveillance analyses (31, 32), is one of the greatest advantages of the cgMLST scheme. Implementing this scheme in the SeqSphere$^+$ software (Ridom GmbH, Münster, Germany) allows the definition and curation of an international and standardized nomenclature, which has successfully been developed for other meaningful pathogens, such as *Staphylococcus aureus*, *Listeria monocytogenes*, and *Enterococcus faecium* (29, 33, 34).

In this study, we generated a powerful typing scheme for *E. faecalis* using the SeqSphere$^+$ software, hence providing an international standardized nomenclature that is suitable for surveillance approaches, outbreak investigations, and phylogenetic analyses.

MATERIALS AND METHODS

Isolate collection. To define the scheme, a total of 146 *E. faecalis* isolates (Table S1) were selected from the strain collection of the German National Reference Centre (NRC) for Staphylococci and Enterococci. The collection comprises isolates from different sources (hospital associated, human colonization, animal, and food) sent to the NRC over a period of 20 years. Hospital-associated strains were isolated from blood (bacteremia or endocarditis), urine, and wounds. Isolates designated human colonization were obtained from stool specimens or rectal swabs as part of patient screening. The animal-associated strains were isolated from dairy cattle, pig, and poultry specimens that included healthy animals, confirmed infections (mastitis), and food (meat). The selection of isolates also included VRE of the most abundant vancomycin-resistant genotypes, *vanA* (*n* = 24) and *vanB* (*n* = 19). For calibration of the novel cgMLST scheme, a subset of closely related *E. faecalis* strains was included. Of these, 27 isolates that were assigned to 10 PFGE types were obtained from an intestinal colonization screening (ICS) program of healthy, nonhospitalized families in Germany. Also, 23 isolates from 4 putative hospital transmission (PHT) events and forming 4 PFGE types (>82% identity) were selected.

Furthermore, two additional isolate collections were compiled to test the newly defined cgMLST scheme (Table S1). First, a collection of 14 *E. faecalis* isolates from blood cultures, representing all *E. faecalis* blood culture isolates sent to the NRC in the years 2015 to 2017, was sequenced and analyzed. Second, a total of 21 environmental samples, isolated from hospital wastewater and the follow-up treatment stages, were analyzed.

DNA extraction, whole-genome sequencing, and *de novo* assembly. Bacterial strains were cultivated overnight in brain heart infusion broth at 37°C. DNA was extracted using a DNeasy blood and tissue kit according to the protocol of the manufacturer (Qiagen, Hilden, Germany). The Qubit double-stranded DNA (dsDNA) high-sensitivity (HS) assay kit (Invitrogen/Thermo Fisher Scientific, Karlsruhe, Germany) was used for DNA quantification. A total of 1 ng of extracted DNA was employed for library preparation using the Nextera XT DNA library prep kit according to the manufacturer's instructions (Illumina, San Diego, CA). Whole-genome sequencing was accomplished using short-read (2 × 300-bp) paired-end sequencing provided by MiSeq ($n = 109$) and HiSeq ($n = 72$) instruments (Illumina, San Diego, CA) as described elsewhere (15). The resulting raw reads were assembled *de novo* using SPAdes (v. 3.9.0, default parameters) (35), which is also integrated into the Linux version of the SeqSphere$^+$ software (v. 5.0.0; Ridom GmbH, Münster, Germany). The assembled sequence data were used for extraction of STs and further downstream analyses.

Single nucleotide polymorphism-based mapping. For comparative purposes and validation of the cgMLST scheme, the sequence data obtained were also used for a single-nucleotide polymorphism (SNP) mapping approach. To this end, raw reads were trimmed using Trimmomatic (v. 0.32; default parameters with sliding window set to 4:15) (36), and resulting paired-end reads were aligned to the reference sequence of *E. faecalis* strain OG1RF (GenBank accession no. NC_017316) using BWA-SW (v. 0.7.13-r1126; default parameters) (37). Subsequent variant calling was performed using VarScan (v. 2.3; default parameters) (38). *E. faecalis* is highly recombinogenic; therefore, SNPs located within a distance of 300 bp or less to each other were excluded using SNPfilter (v. 3.2.3; exclusion distance [d] = 300) (39). Maximum likelihood phylogenetic trees were calculated on the basis of retained SNPs using PhyML with a general time reversible (GTR) nucleotide model (v. 3.0; bootstrap, 1,000 permutations) (40). Phylogenetic trees were visualized by applying iTOL (v. 4.2.3; https://itol.embl.de/) (41).

Target scheme definition. The finished and publicly available genome of *E. faecalis* OG1RF (version 1, GenBank accession no. NC_017316.1) was selected as the "seed genome" for the cgMLST scheme definition. Using the cgMLST Target Definer tool (v. 1.5 with default parameters) of the Ridom SeqSphere$^+$ v. 4.0.9 software, a rapid local *ad hoc* cgMLST scheme was defined that contained all genes of the reference genome that were not homologous, did not contain internal stop codons, and did not overlap other genes. All other 11 publicly accessible and finished *E. faecalis* genome sequences (as of 12 May 2017) (Table S1), together with the 146-shotgun genome collection of this study (see above), were selected as query genomes to generate a cgMLST task template applicable to *E. faecalis* strains of various origins. Only targets from the *ad hoc* scheme that were present in ≥95% of all query genomes were accepted as cgMLST targets. This parameter was chosen due to the genetic plasticity and the diverse population structure of *E. faecalis* and to ensure a typing scheme suitable for isolates of various origins. The resulting compilation of target genes (Data Set S1) was defined as the core genome genes and used for the subsequent typing scheme.

Evaluation of the cgMLST target gene set. All publicly available genome sequences from the NCBI and NCBI Sequence Read Archive (SRA) databases (as of 30 August 2017) were used to evaluate the cgMLST scheme. In total, 526 NCBI genomes were downloaded and analyzed in SeqSphere$^+$. To ensure a defined and quality-controlled data set for the evaluation process, a thorough manual filtering process was performed. First, all genomes used to create the cgMLST task template were removed from the evaluation data set ($n = 514$ after filtering). Next, all data sets sequenced with Ion Torrent or 454 technology were excluded because of their increased tendency to contain homopolymer errors ($n = 502$). Afterwards, and in order to ensure sufficient quality of the sequence data used, all data records for which no MLST could be assigned by SeqSphere$^+$ were filtered out ($n = 488$). Finally, the data sets were searched for duplicates, since prominent or culture collection strains are often overrepresented and could thus induce a bias in the evaluation data set ($n = 481$). Genome sequences from the SRA were automatically filtered with SeqSphere$^+$ by (i) Illumina data, (ii) SRA replicate filtering (technical replicates of one SRA sample were identified and the biggest data set kept), and (iii) strain name filtering, to detect strains submitted more than once, e.g., by different submitters (the largest data set was selected). After filtering, 984 of 2,093 genome sequences met these criteria. One single-genome sequence was unavailable, so a total of 983 SRA sequences were finally downloaded for scheme evaluation. All data sets were assembled with SPAdes (v. 3.9.0 implemented in SeqSphere$^+$ with default settings ["–careful" option turned off]) and analyzed with SeqSphere$^+$. Most of the manually processed filter criteria used for the NCBI data were implemented in the SRA download and automatically processed. Therefore, only SRA data duplicates of the seed and query genomes ($n = 978$), as well as low-quality data, i.e., data sets to which no MLST could be assigned, had to be removed by manual filtering, leaving 849 data sets of the NCBI-SRA collection. A final manual filtering step omitted 330 genome data duplicates derived from merging the SRA and NCBI sequences (NCBI data were kept while SRA data were removed). Finally, 1,000 data sets were retained for cgMLST evaluation.

cgMLST-based analysis. The environmental strains, NRC and NCBI strains, and all isolates used to define the novel cgMLST were analyzed using the cgMLST scheme to validate the applicability of the scheme and review the collection for its representativeness of the *E. faecalis* population structure. The software defined individual allelic profiles at the strain level. These profiles were used to construct minimum spanning trees (MST) in SeqSphere$^+$ for the entire strain collection by pairwise ignoring missing values. Furthermore, phylogenetic trees were calculated using the neighbor-joining algorithm with default parameters implemented in the Ridom SeqSphere$^+$ software. Trees were visualized using iTOL (v. 4.2.3) (41).

Determination of cluster distance. To determine allelic distances between closely related isolates, the calibration subsets of isolates obtained from the ICS program ($n = 27$) and from PHT events ($n = 23$)

were used. Detailed knowledge about the epidemiological background of these subsets further strengthened the selection chosen for calibration. The threshold for cluster distance determination was set based on the results of the cgMLST-based MST using the allelic differences between highly identical isolates compared with allelic differences from or between nonrelated isolates. For validation purposes, the threshold derived was compared with results of the SNP-based mapping analyses.

Comparison of phylogenetic trees. The resulting phylogenetic trees from the SNP- and cgMLST-based approaches to the collection of 146 isolates were used to compare the concordance of the two methods. The phylogenies were checked for similarities and differences regarding isolate clustering. In addition, using normalized absolute differences of patristic distances of isolate pairs obtained by the respective phylogenetic analyses, an isolate-by-isolate comparison was performed. It was necessary to multiply the difference values by 1×10^9, followed by taking the logarithm (basis = 2), to generate distance values suitable for comparing the two approaches. These calculations resulted in values that were used as a measure of the grade of concordance, which then was visualized by heat maps using iTOL (v. 4.2.3) (41). In doing so, the logarithm compressed the data distribution. To mitigate this effect in the visualization, the median of the values was calculated and used as the mean value of the color scale.

FastTree analyses with available sequence types. A maximum-likelihood phylogenetic tree was generated by FastTree 2 (v. 2.1.7, Jukes-Cantor+CAT model) using aligned concatenated sequence data of 852 MLST sequence types (as of 25 January 2018) obtained from PubMLST (www.pubmlst.org) (42, 43). The resulting phylogenetic tree was visualized by applying iTOL (v. 4.2.3) (41).

Accession number(s). Raw reads of all sequenced *E. faecalis* isolates of this study ($n = 181$) were submitted to the Sequence Read Archive database of the National Center for Biotechnology Information and are available under accession no. SRP156712.

RESULTS

Diversity and suitability of the strain collection for scheme definition. MLST analyses revealed 41 different STs present in the strain collection of 146 isolates, 7 of which were new STs submitted to the PubMLST database (42). Considering the diverse origin of the isolates (hospital associated, human colonization, animal, and food) and the sampling period of more than 20 years, the collection provided a deep insight into the population of *E. faecalis*.

To visualize the genetic relation between isolates of the collection, a phylogenetic tree based on MLST genes was calculated using the Ridom SeqSphere+ software (Fig. 1). The MLST tree was annotated with available metadata of all isolates, in particular, ST, origin, vancomycin resistance, and year of isolation. In due consideration of the importance of the human- and clinical-associated lineage ST6, which also included most of the VRE, this genotype was more frequently represented in the collection ($n = 47$, ~32%).

To gain an insight into the representativeness of the strain collection in the context of all available STs, a maximum-likelihood phylogenetic tree based on sequences of 852 STs was calculated using FastTree 2 (Fig. 2). The resulting tree showed clustering of available sequence types, with each single leaf representing one ST. As can be seen in Fig. 2, the selected strains were distributed over the entire tree and covered almost every branch, indicating that the strain collection reflects an extensive and diverse selection of the *E. faecalis* ST-based population structure (Fig. 2B). Only an outlier group of five different STs was not represented in the isolate collection used to define the cgMLST scheme (Fig. 2A).

Development and evaluation of the *E. faecalis* cgMLST scheme. Based on the *E. faecalis* strain OG1RF reference genome (2,636 genes with coding DNA sequences), a preliminary reference task template was generated that retained a total of 2,385 target genes after thorough filtering. Using all 11 finished genome sequences from NCBI (as of 15 January 2017) and the 146 genome sequences of our strain collection as penetration query genomes, a cgMLST scheme of 1,972 target genes was defined, covering 67.9% of the seed genome (Data set S1). All finally assigned cgMLST targets were present in ≥95% of all query genomes.

For evaluation purposes, 1,000 genome sequences from the NCBI and SRA databases (as of 30 August 2017) were downloaded, thoroughly filtered, and finally analyzed using the *E. faecalis* cgMLST scheme. At least 95% good cgMLST target genes were found in more than 98.6% of the genome sequences (mean, 99.2% good target genes). In addition, the 146 query genome sequences of the isolate collection were also used to analyze the performance of the newly defined cgMLST scheme. Query genome
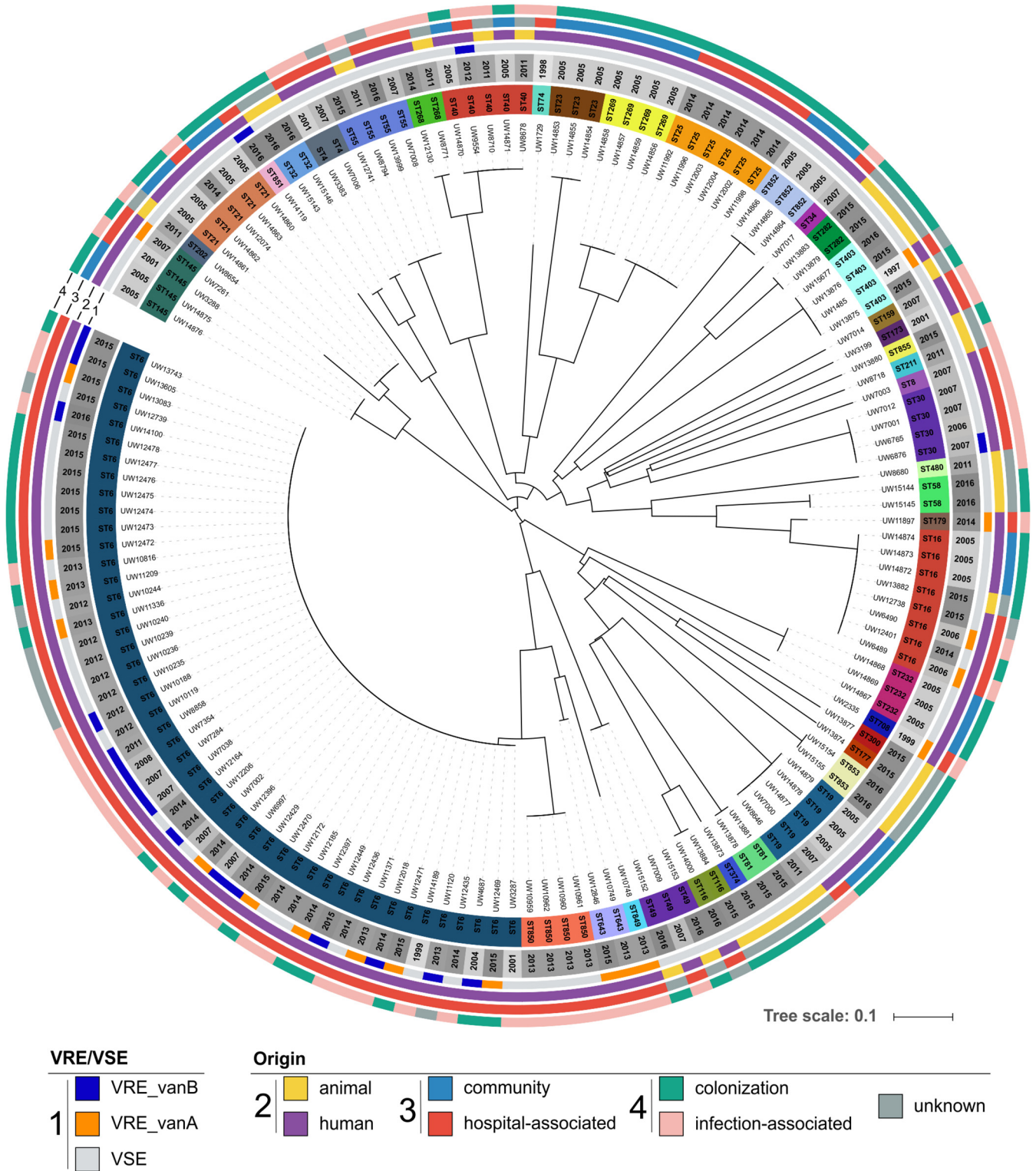
FIG 1 Isolate collection for establishing the *E. faecalis* cgMLST scheme (*n* = 146). The phylogenetic tree is based on MLST analyses and was calculated with SeqSphere⁺. Sequence types (STs) were added and color-coded in the inner circle. Year of strain isolation was added and shaded as a gray circle (light gray to dark gray). Vancomycin resistance and the *van* genotype were also added as color-coded circles (see key, "VRE/VSE"). In the three outer circles (2 to 4), the origin of isolation was depicted (see key, "Origin"). Circle 2 depicts whether an isolate is of human (purple) or animal (yellow) origin. Circle 3 represents the origin of human isolates (blue, community; red, hospital associated). The outer circle, 4, shows whether an isolate was obtained from colonization (green) or was infection associated (pink). For some isolates, the data were not valid or were unknown (dark gray). Visualization was realized using iTOL.
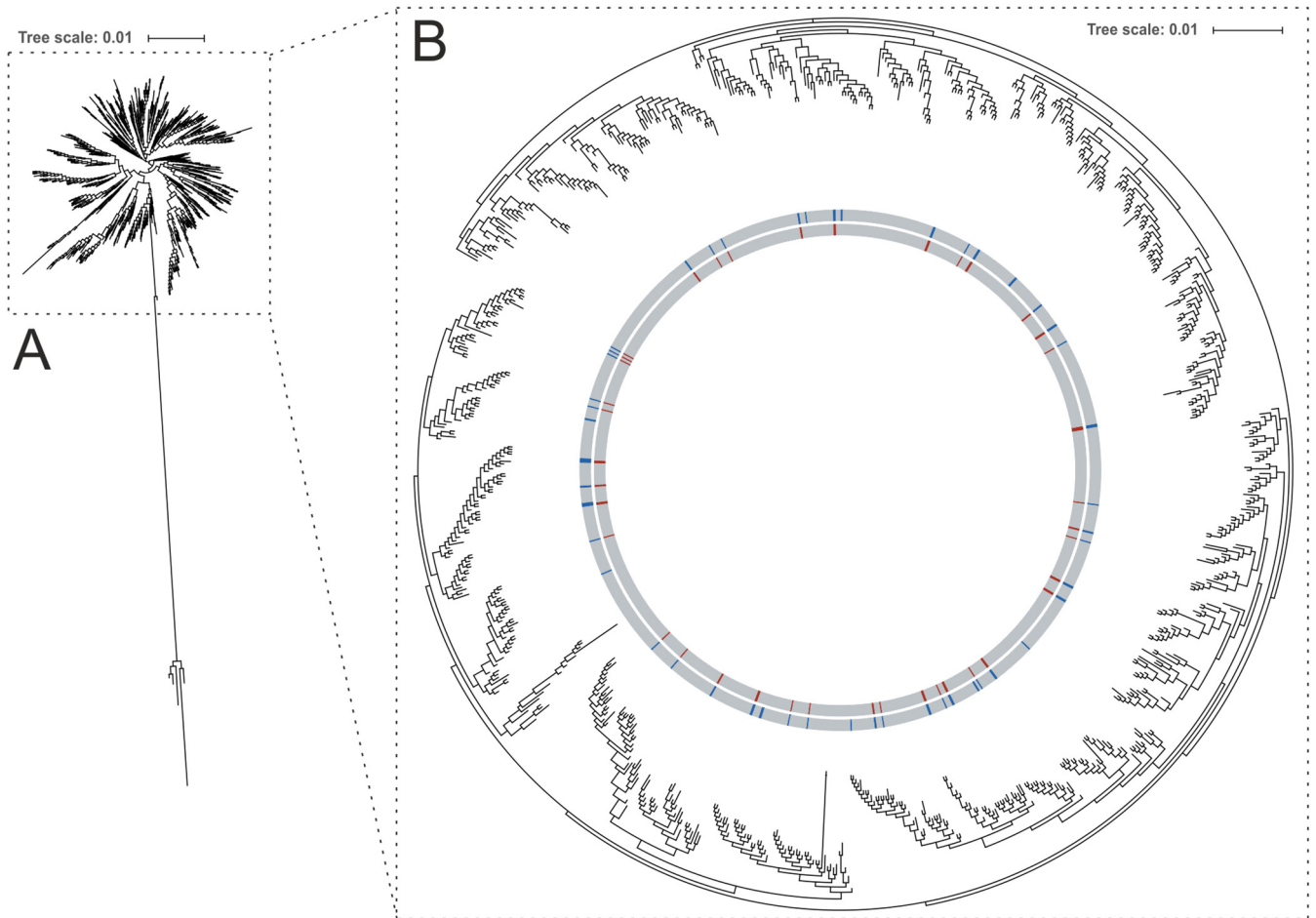
FIG 2 The phylogenetic tree of 852 sequence types (STs) represents the distribution of STs used for cgMLST scheme development within the global ST population of *E. faecalis* and was generated using FastTree 2. (A) A radial overview and (B) a circular presentation of the main part of the tree are shown. STs represented in the strain collection for scheme definition are color coded in red. Blue color codes represent all STs available for this study. Visualization was realized using iTOL.

analysis resulted in a mean of 99.6% good targets detected (ranging from 96.4% to 100%) (Table S1; "% good cgMLST targets" column).

To determine cluster distances, a subset of 50 isolates with detailed epidemiological data (ICS, PHT) was used. The ICS isolates belonged to 10 distinct groups determined by PFGE analyses and comprised 9 different STs (Fig. S1). Relatedness was further supported by the results from the SNP-based analysis, as these groups exhibit 0 to 1 SNP differences (Data Set S2). The isolates from PHT also belonged to four distinct PFGE groups, comprised three different STs, and demonstrated a close relationship based on SNP analyses (0 to 7 SNP differences; Data Set S2). Analysis of the 50 calibration isolates using the novel *E. faecalis* cgMLST scheme confirmed the close relationship of those strains (Fig. S2A). Based on these results, and considering highly related and nonrelated members of the strain collection, we defined a difference of seven alleles as the threshold for distinct clusters.

Concordance of the cgMLST scheme with SNP-based phylogenetic analyses. To evaluate the newly defined cgMLST scheme, the results obtained were compared with an SNP-based mapping approach. A total of 2,200 SNP positions (recombinatoric events excluded by distance-based filtering) were considered for tree calculation. The resulting maximum likelihood phylogenetic tree had 147 tips (including reference genome) with 292 nodes, and it demonstrated that the isolates differed from each other by 0 to 213 SNPs (Fig. S3A; Data Set S2). The cgMLST-based phylogenetic tree consisted of 146 tips

and 291 nodes, roughly indicating comparable outcomes (Fig. S3B). Both methods for phylogenetic tree calculation showed concordant clustering according to ST grouping and that of isolate clusters (>2 isolates) comprising several STs (Fig. S3).

For visualization and as a measure of concordance, a heat map was created, which was calculated based on the normalized absolute patristic distances for each particular isolate obtained by the two different approaches (Fig. S4). The heat map comparison showed a good concordance for closely related strains, such as the ST6 subpopulation, and for the groups of the ICS program (Fig. S4, visualized by blue boxes). The more distantly related isolates performed differently with the two approaches (Fig. S4, visualized by orange boxes).

Application of the novel cgMLST scheme to describe the *E. faecalis* population structure. The cgMLST scheme was also tested to investigate the relatedness of strains putatively involved in hospital outbreak events and to obtain further insights into *E. faecalis* population structure. Over 95% good cgMLST target genes were found in all isolate sequences of the entire strain collection, including those that were not used to define the cgMLST scheme (mean, 99.7% good target genes). The allocation of complex types (CTs) showed that the collection was composed of 97 different CTs (Fig. 3). To further visualize the results, an MST was calculated (Fig. S2). The isolates that were chosen for CT calibration were clearly represented by distinct clusters in the MST analysis (Fig. S2A). Due to the epidemiological data and previous PFGE analyses, PHT group 1 was identified, and indeed showed a maximum difference of six alleles in the cgMLST analysis. The isolates belonged to ST25 and CT722. The NCBI isolate *E. faecalis* LD33, which is also ST25 but exhibits CT237, was unequivocally separated from PHT group 1 by 490 alleles. Furthermore, isolates belonging to PHT groups 2 and 3 displayed identical CTs within each cluster and thus also proved suitable for complex calibration (0- to 2-allele difference). In contrast, PHT group 4 comprises eight ST6 isolates that were also demonstrated to be closely related in previous PFGE analyses (>85% identity; Fig. S1) and in SNP-based examination (0 to 11 SNPs; Data Set S2) but yielded four different CTs (0 to 15-allele difference; Fig. S2B).

The calibration collection also included 27 isolates sampled as part of an ICS program of healthy families in Germany (Fig. S2A). Previous PFGE analyses revealed that these isolates belonged to 10 distinct groups, with each screened family creating a distinct group (data not shown). The cgMLST analyses (>99.6% good target genes) generated results concordant with PFGE typing. The ICS groups generated no separate cluster of commensal isolates within the population, and instead these isolates were widely distributed over the MST (Fig. S2). Altogether, the ICS groups were suitable for calibration, as they differed by at most two alleles from each other within one group, but by over 29 alleles in relation to unrelated isolates harboring identical STs.

In addition, a selection of isolates that were not used to define the cgMLST scheme was analyzed. These isolates included 14 clinical isolates and also 21 environmental samples of *E. faecalis*. With a mean of 99.4% good target genes, all environmental isolates were successfully analyzed. Six isolates belonged to ST6 and were obtained directly from clinical sewage. Concordantly, those isolates clustered together with ST6 isolates of the strain collection in MST analyses (Fig. S2B). Nevertheless, they constituted a distinct cluster when considering the corresponding distance matrix (>27-allele difference from other ST6 isolates) and exhibited CT744. The other environmental isolates were distributed over the tree, harbored unique CTs, and showed no distinct clustering depending on the sampling site (Fig. S2A).

The 14 clinical isolates that were obtained from blood culture and sent to the NRC in the years 2015 to 2017 yielded 11 different STs and only unique CTs (mean, 99.4% good target genes). Four isolates were assigned to ST6 and also clustered within the ST6 group, but in contrast to the sewage, ST6 isolates did not separate into a distinct clade (Fig. S2B).
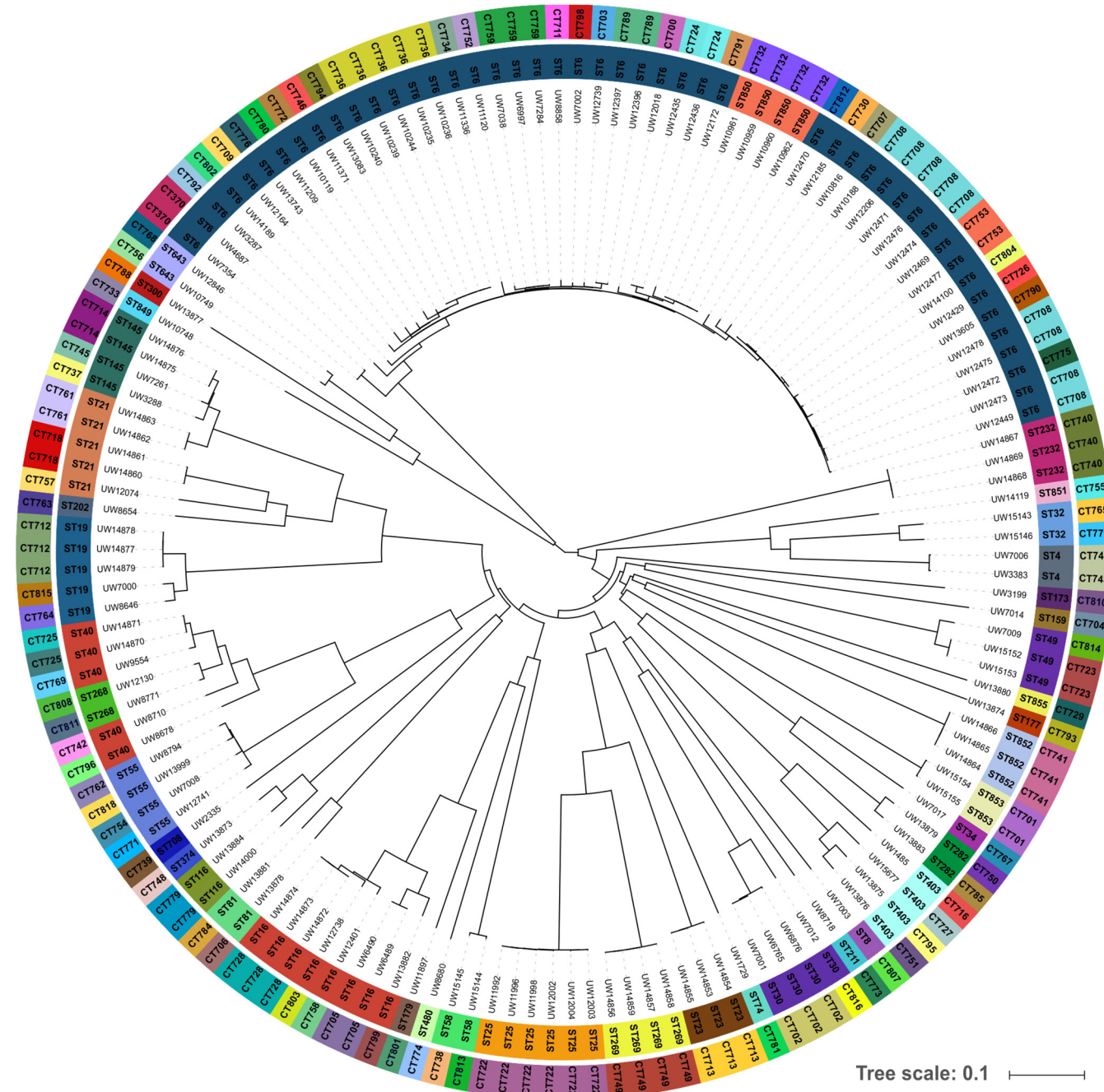
FIG 3 Isolate collection for establishing the *E. faecalis* cgMLST scheme (*n* = 146). The phylogenetic tree is based on cgMLST analyses and was calculated with SeqSphere+. Available data for sequence types (STs; inner circle) and complex types (CTs; outer circle) are depicted as colored circles. The collection comprised 41 different STs and 97 different CTs. Visualization was realized using iTOL.

DISCUSSION

In this study, we developed and evaluated a robust typing scheme for analyzing *E. faecalis*, an opportunistic pathogen of various origins. The scheme was implemented in the Ridom SeqSphere+ software suite to generate highly concordant and comparable results and to provide a standardized nomenclature for the scientific community.

The *E. faecalis* strain OG1RF was used as the seed genome to define the cgMLST targets. Based on previous investigations of comparative genomics, the OG1RF strain represents a rather conserved *E. faecalis* genome due to the lack of mobile genetic elements (44). In contrast, and thus less suitably for a cgMLST definition, the fully

sequenced *E. faecalis* strain V583 was demonstrated to contain up to 25% mobile genetic elements within its genome (18).

The novel cgMLST scheme proved to be a powerful tool for analyzing *E. faecalis* isolates, including those of various sampling backgrounds, such as hospital-associated strains and strains obtained from general screenings in outpatient settings (colonization), but also for strains from animals, food, and the environment. Furthermore, the results of the evaluation with genome sequences from the NCBI and SRA databases highlight the robustness of the selected target genes and thus the suitability of the scheme to investigate the relatedness of *E. faecalis* isolates based on core genome data.

The results have also provided a detailed overview that enhances our current and common understanding of the *E. faecalis* population structure. When the classical MLST scheme for *E. faecalis* was published, the first insights into the population structure of this opportunistic pathogen became available (23). The authors described correlating results obtained by MLST and PFGE, but they also mentioned slight deviations between the two typing methods and additionally noted the influence of mobile genetic elements and high rates of recombination events. In accordance with the results of the cgMLST analyses in the present study, a dispersion of single genotypes over various origins and sampling sites was reported. However, a few lineages (e.g., ST6) showed adaptation to the hospital environment, which is especially known for *E. faecium* clinical isolates (21, 45–47).

The 14 isolates from blood cultures that were received by the NRC between 2015 and 2017 are represented by 11 different sequence types and 14 different complex types. This again supports the hypothesis that there is no distinct lineage responsible for human infections. Nonetheless, four of the isolates belonged to ST6, which suggests that at least some STs might primarily be human associated (21, 23). These results are in accordance with ST6, which is detected with a high prevalence in hospitalized patients (23, 48), as a human-associated lineage. Tedim et al. attributed the high prevalence of ST6 (and other highly abundant STs) to their high potential for adaptation (45). To create a robust scheme for *E. faecalis* outbreak analysis, we therefore included a high number ($n = 47$; 32%) of ST6 isolates in our strain collection used for scheme definition. The ST6 isolates incorporated into this study were collected over 18 years and originated exclusively from humans, including from the most common clinical presentations, such as bacteremia and endocarditis. In total, the ST6 strains analyzed comprised 57 isolates that yielded 34 different complex types. As was anticipated prior to setting up the cgMLST scheme, subdifferentiation was able to reveal multiple CTs within a certain ST, and thus might be suitable for tracing or excluding transmission events.

One considerable advantage of cgMLST, especially for genotyping purposes, is the focus on alleles instead of SNPs. While a missing gene or homologous recombination events can result in numerous SNPs, only entire allele changes are taken into account by cgMLST analyses (31, 49). In addition, the threshold of CT definition mitigates possible typing variations caused by these events. Indeed, cgMLST provides a robust and expandable database for allelic profiles, and it has already been applied for outbreak analyses of various pathogenic bacteria, such as *Enterococcus faecium*, *Staphylococcus aureus*, *Listeria monocytogenes*, and *Clostridium difficile* (29, 33, 34, 50).

Besides cgMLST, which is superior in its resolution compared to classical typing methods, reference-based bioinformatics approaches are widely used for more detailed analysis of strain relatedness and for phylogenetic purposes (51). Such approaches provide results in high resolution for investigating closely related individuals, as in outbreaks mostly restricted to a local setting (e.g., one ward, one hospital, or one country) (52). In contrast, phylogenetic investigations on a "global" scale or long-term studies often lack a reference suitable for the entire set of strains to be analyzed (53). Although programs such as refRank will select for the best reference for a given data set, it is always a trade-off with respect to the genomic variety of the strain collection

(39). In addition, SNP-based approaches must be adjusted for recombination events, which would otherwise impact phylogenetic interpretations (49, 54).

Assessing the phylogenetic trees of the two approaches, the phylogenies seemed to be conformable and reproducible, as described before for comparative analyses of *L. monocytogenes* (55). Also, our mathematical approach to comparing the concordance of patristic distances obtained by SNP-based and cgMLST analyses demonstrated that a high concordance could be reached for closely related isolates. Isolates without a close relationship showed less concordant results. Taking into consideration that the SNP-based phylogenetic approach also includes intergenic regions, while cgMLST does not, the low concordance observed for distantly related isolates is expected.

We also analyzed and inversely examined the concordance of the cgMLST and SNP-based approaches using our strain collection. For instance, isolates of one of the putative hospital transmission groups (PHT group 2) were proven to be closely related both by SNP-based mapping and by cgMLST data. Analyses of the PHT group 2 and 3 strains, which originated from the same hospital and ward, were isolated 6 months apart and harbored different STs and CTs, a finding that was also supported by the SNP-based approach (7 SNPs). For isolates of PHT group 4, the cgMLST analyses showed that these strains belonged to three different CTs, revealing that this group was not built by the spread of a single clone. These findings again indicate the general applicability of cgMLST. Furthermore, it highlights the need to select a suitable reference genome for downstream analyses of variant-calling investigations in order to produce reliable results, by applying the most reasonable method for data evaluation at a given time and for a certain strain collection.

In conclusion, we successfully defined a cgMLST scheme for *E. faecalis*. Our analyses showed that cgMLST was suitable for analyzing closely related strains from putative hospital transmission events, as well as for analyzing and portraying the general population structure of this important opportunistic pathogen.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/JCM .01686-18.

SUPPLEMENTAL FILE 1, XLS file, 0.3 MB.
SUPPLEMENTAL FILE 2, XLS file, 0.2 MB.
SUPPLEMENTAL FILE 3, PDF file, 0.8 MB.
SUPPLEMENTAL FILE 4, PDF file, 1.6 MB.
SUPPLEMENTAL FILE 5, PDF file, 1.5 MB.
SUPPLEMENTAL FILE 6, PDF file, 1.7 MB.
SUPPLEMENTAL FILE 7, PDF file, 0.2 MB.

REFERENCES

1. Abriouel H, Omar NB, Molinos AC, López RL, Grande MJ, Martínez-Viedma P, Ortega E, Cañamero MM, Galvez A. 2008. Comparative analysis of genetic diversity and incidence of virulence factors and antibiotic resistance among enterococcal populations from raw fruit and vegetable foods, water and soil, and clinical samples. Int J Food Microbiol 123:38–49. https://doi.org/10.1016/j.ijfoodmicro.2007.11.067.

2. Franz CMAP, Huch M, Abriouel H, Holzapfel W, Gálvez A. 2011. Enterococci as probiotics and their implications in food safety. Int J Food Microbiol 151:125–140. https://doi.org/10.1016/j.ijfoodmicro.2011.08.014.

3. Giraffa G. 2003. Functionality of enterococci in dairy products. Int J Food Microbiol 88:215–222.

4. Domann E, Hain T, Ghai R, Billion A, Kuenne C, Zimmermann K, Chakraborty T. 2007. Comparative genomic analysis for the presence of potential enterococcal virulence factors in the probiotic *Enterococcus faecalis* strain Symbioflor 1. Int J Med Microbiol 297:533–539. https://doi.org/10.1016/j.ijmm.2007.02.008.

5. Buhnik-Rosenblau K, Matsko-Efimov V, Danin-Poleg Y, Franz CM, Klein G, Kashi Y. 2013. Biodiversity of *Enterococcus faecalis* based on genomic typing. Int J Food Microbiol 165:27–34. https://doi.org/10.1016/j.ijfoodmicro.2013.04.009.

6. Larsen J, Schonheyder HC, Lester CH, Olsen SS, Porsbo LJ, Garcia-Migura L, Jensen LB, Bisgaard M, Hammerum AM. 2010. Porcine-origin gentamicin-resistant *Enterococcus faecalis* in humans, Denmark. Emerg Infect Dis 16:682–684. https://doi.org/10.3201/eid1604.090500.

7. Abat C, Huart M, Garcia V, Dubourg G, Raoult D. 2016. *Enterococcus faecalis* urinary-tract infections: do they have a zoonotic origin? J Infect 73:305–313. https://doi.org/10.1016/j.jinf.2016.07.012.

8. Werner G, Fleige C, Neumann B, Bender JK, Layer F, Klare I. 2015. Evaluation of DiversiLab(R), MLST and PFGE typing for discriminating clinical *Enterococcus faecium* isolates. J Microbiol Methods 118:81–84. https://doi.org/10.1016/j.mimet.2015.08.019.

9. Bohme H, Konigsmark C, Klare I, Zischka M, Werner G. 2012. Cross-transmission rates of enterococcal isolates among newborns in a neonatal intensive care unit. Pediatr Rep 4:e15. https://doi.org/10.4081/pr.2012.e15.

10. Werner G, Fleige C, Fessler AT, Timke M, Kostrzewa M, Zischka M, Peters T, Kaspar H, Schwarz S. 2012. Improved identification including MALDI-TOF mass spectrometry analysis of group D streptococci from bovine mastitis and subsequent molecular characterization of corresponding *Enterococcus faecalis* and *Enterococcus faecium* isolates. Vet Microbiol 160:162–169. https://doi.org/10.1016/j.vetmic.2012.05.019.

11. Wendt C, Wiesenthal B, Dietz E, Ruden H. 1998. Survival of vancomycin-resistant and vancomycin-susceptible enterococci on dry surfaces. J Clin Microbiol 36:3734–3736.

12. Wille I, Mayr A, Kreidl P, Bruhwasser C, Hinterberger G, Fritz A, Posch W, Fuchs S, Obwegeser A, Orth-Holler D, Lass-Florl C. 2017. A cross sectional point prevalence survey to study the environmental contamination of nosocomial pathogens in intensive care units under real life conditions. J Hosp Infect https://doi.org/10.1016/j.jhin.2017.09.019.

13. Murdoch DR, Corey GR, Hoen B, Miro JM, Fowler VG, Jr, Bayer AS, Karchmer AW, Olaison L, Pappas PA, Moreillon P, Chambers ST, Chu VH, Falco V, Holland DJ, Jones P, Klein JL, Raymond NJ, Read KM, Tripodi MF, Utili R, Wang A, Woods CW, Cabell CH. 2009. Clinical presentation, etiology, and outcome of infective endocarditis in the 21st century: the International Collaboration on Endocarditis-Prospective Cohort Study. Arch Intern Med 169:463–473. https://doi.org/10.1001/archinternmed.2008.603.

14. Hegstad K, Mikalsen T, Coque TM, Werner G, Sundsfjord A. 2010. Mobile genetic elements and their contribution to the emergence of antimicrobial resistant *Enterococcus faecalis* and *Enterococcus faecium*. Clin Microbiol Infect 16:541–554. https://doi.org/10.1111/j.1469-0691.2010.03226.x.

15. Bender JK, Kalmbach A, Fleige C, Klare I, Fuchs S, Werner G. 2016. Population structure and acquisition of the *vanB* resistance determinant in German clinical isolates of *Enterococcus faecium* ST192. Sci Rep 6:21847. https://doi.org/10.1038/srep21847.

16. Arias CA, Murray BE. 2009. Antibiotic-resistant bugs in the 21st century—a clinical super-challenge. N Engl J Med 360:439–443. https://doi.org/10.1056/NEJMp0804651.

17. Hidron AI, Edwards JR, Patel J, Horan TC, Sievert DM, Pollock DA, Fridkin SK, National Healthcare Safety Network Team, Participating National Healthcare Safety Network Facilities. 2008. NHSN annual update: antimicrobial-resistant pathogens associated with healthcare-associated infections: annual summary of data reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2006–2007. Infect Control Hosp Epidemiol 29:996–1011. https://doi.org/10.1086/591861.

18. Paulsen IT, Banerjei L, Myers GS, Nelson KE, Seshadri R, Read TD, Fouts DE, Eisen JA, Gill SR, Heidelberg JF, Tettelin H, Dodson RJ, Umayam L, Brinkac L, Beanan M, Daugherty S, DeBoy RT, Durkin S, Kolonay J, Madupu R, Nelson W, Vamathevan J, Tran B, Upton J, Hansen T, Shetty J, Khouri H, Utterback T, Radune D, Ketchum KA, Dougherty BA, Fraser CM. 2003. Role of mobile DNA in the evolution of vancomycin-resistant *Enterococcus faecalis*. Science 299:2071–2074. https://doi.org/10.1126/science.1080613.

19. Leavis HL, Bonten MJ, Willems RJ. 2006. Identification of high-risk enterococcal clonal complexes: global dispersion and antibiotic resistance. Curr Opin Microbiol 9:454–460. https://doi.org/10.1016/j.mib.2006.07.001.

20. Freitas AR, Novais C, Ruiz-Garbajosa P, Coque TM, Peixe L. 2009. Clonal expansion within clonal complex 2 and spread of vancomycin-resistant plasmids among different genetic lineages of *Enterococcus faecalis* from Portugal. J Antimicrob Chemother 63:1104–1111. https://doi.org/10.1093/jac/dkp103.

21. Kawalec M, Pietras Z, Daniłowicz E, Jakubczak A, Gniadkowski M, Hryniewicz W, Willems RJL. 2007. Clonal structure of *Enterococcus faecalis* isolated from Polish hospitals: characterization of epidemic clones. J Clin Microbiol 45:147–153. https://doi.org/10.1128/JCM.01704-06.

22. Solheim M, Aakra A, Snipen LG, Brede DA, Nes IF. 2009. Comparative genomics of *Enterococcus faecalis* from healthy Norwegian infants. BMC Genomics 10:194. https://doi.org/10.1186/1471-2164-10-194.

23. Ruiz-Garbajosa P, Bonten MJ, Robinson DA, Top J, Nallapareddy SR, Torres C, Coque TM, Canton R, Baquero F, Murray BE, del Campo R, Willems RJ. 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. https://doi.org/10.1128/JCM.02596-05.

24. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG. 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.

25. Pinholt M, Larner-Svensson H, Littauer P, Moser CE, Pedersen M, Lemming LE, Ejlertsen T, Sondergaard TS, Holzknecht BJ, Justesen US, Dzajic E, Olsen SS, Nielsen JB, Worning P, Hammerum AM, Westh H, Jakobsen L. 2015. Multiple hospital outbreaks of *vanA Enterococcus faecium* in Denmark, 2012–13, investigated by WGS, MLST and PFGE. J Antimicrob Chemother 70:2474–2482. https://doi.org/10.1093/jac/dkv142.

26. Salipante SJ, SenGupta DJ, Cummings LA, Land TA, Hoogestraat DR, Cookson BT. 2015. Application of whole-genome sequencing for bacterial strain typing in molecular epidemiology. J Clin Microbiol 53:1072–1079. https://doi.org/10.1128/JCM.03385-14.

27. Didelot X, Bowden R, Wilson DJ, Peto TEA, Crook DW. 2012. Transforming clinical microbiology with bacterial genome sequencing. Nat Rev Genet 13:601–612. https://doi.org/10.1038/nrg3226.

28. Reuter S, Ellington MJ, Cartwright EJ, Koser CU, Torok ME, Gouliouris T, Harris SR, Brown NM, Holden MT, Quail M, Parkhill J, Smith GP, Bentley SD, Peacock SJ. 2013. Rapid bacterial whole-genome sequencing to enhance diagnostic and public health microbiology. JAMA Intern Med 173:1397–1404. https://doi.org/10.1001/jamainternmed.2013.7734.

29. 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 RJ. 2015. Core genome multilocus sequence typing scheme for high-resolution typing of *Enterococcus faecium*. J Clin Microbiol 53:3788–3797. https://doi.org/10.1128/JCM.01946-15.

30. Mellmann A, Harmsen D, Cummings CA, Zentz EB, Leopold SR, Rico A, Prior K, Szczepanowski R, Ji Y, Zhang W, McLaughlin SF, Henkhaus JK, Leopold B, Bielaszewska M, Prager R, Brzoska PM, Moore RL, Guenther S, Rothberg JM, Karch H. 2011. Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. PLoS One 6:e22751. https://doi.org/10.1371/journal.pone.0022751.

31. Higgins PG, Prior K, Harmsen D, Seifert H. 2017. Development and evaluation of a core genome multilocus typing scheme for whole-genome sequence-based typing of *Acinetobacter baumannii*. PLoS One 12:e0179228. https://doi.org/10.1371/journal.pone.0179228.

32. Kohl TA, Diel R, Harmsen D, Rothganger J, Walter KM, Merker M, Weniger T, Niemann S. 2014. Whole-genome-based *Mycobacterium tuberculosis* surveillance: a standardized, portable, and expandable approach. J Clin Microbiol 52:2479–2486. https://doi.org/10.1128/JCM.00567-14.

33. Ruppitsch W, Pietzka A, Prior K, Bletz S, Fernandez HL, Allerberger F, Harmsen D, Mellmann A. 2015. Defining and evaluating a core genome multilocus sequence typing scheme for whole-genome sequence-based typing of *Listeria monocytogenes*. J Clin Microbiol 53:2869–2876. https://doi.org/10.1128/JCM.01193-15.

34. Leopold SR, Goering RV, Witten A, Harmsen D, Mellmann A. 2014. Bacterial whole-genome sequencing revisited: portable, scalable, and standardized analysis for typing and detection of virulence and antibiotic resistance genes. J Clin Microbiol 52:2365–2370. https://doi.org/10.1128/JCM.00262-14.

35. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.

36. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

37. Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589–595. https://doi.org/10.1093/bioinformatics/btp698.

38. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 22:568–576. https://doi.org/10.1101/gr.129684.111.

39. Becker L, Fuchs S, Pfeifer Y, Semmler T, Eckmanns T, Korr G, Sissolak D, Friedrichs M, Zill E, Tung M-L, Dohle C, Kaase M, Gatermann S, Rüssmann H, Steglich M, Haller S, Werner G. 2018. Whole genome sequence analysis of CTX-M-15 producing *Klebsiella isolates* allowed dissecting a polyclonal outbreak scenario. Front Microbiol 9:322. https://doi.org/10.3389/fmicb.2018.00322.

40. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59:307–321. https://doi.org/10.1093/sysbio/syq010.

41. Letunic I, Bork P. 2016. Interactive Tree of Life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res 44:W242–W245. https://doi.org/10.1093/nar/gkw290.

42. Jolley KA, Maiden MC. 2010. BIGSdb: scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics 11:595. https://doi.org/10.1186/1471-2105-11-595.

43. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490. https://doi.org/10.1371/journal.pone.0009490.

44. Palmer KL, Godfrey P, Griggs A, Kos VN, Zucker J, Desjardins C, Cerqueira G, Gevers D, Walker S, Wortman J, Feldgarden M, Haas B, Birren B, Gilmore MS. 2012. Comparative genomics of enterococci: variation in *Enterococcus faecalis*, clade structure in *E. faecium*, and defining characteristics of *E. gallinarum* and *E. casseliflavus*. mBio 3:e00318-11. https://doi.org/10.1128/mBio.00318-11.

45. Tedim AP, Ruiz-Garbajosa P, Corander J, Rodriguez CM, Canton R, Willems RJ, Baquero F, Coque TM. 2015. Population biology of intestinal enterococcus isolates from hospitalized and nonhospitalized individuals in different age groups. Appl Environ Microbiol 81:1820–1831. https://doi.org/10.1128/AEM.03661-14.

46. Willems RJ, Hanage WP, Bessen DE, Feil EJ. 2011. Population biology of Gram-positive pathogens: high-risk clones for dissemination of antibiotic resistance. FEMS Microbiol Rev 35:872–900. https://doi.org/10.1111/j.1574-6976.2011.00284.x.

47. Lebreton F, van Schaik W, McGuire AM, Godfrey P, Griggs A, Mazumdar V, Corander J, Cheng L, Saif S, Young S, Zeng Q, Wortman J, Birren B, Willems RJ, Earl AM, Gilmore MS. 2013. Emergence of epidemic multidrug-resistant *Enterococcus faecium* from animal and commensal strains. mBio 4:e00534-13. https://doi.org/10.1128/mBio.00534-13.

48. Bello Gonzalez TDJ, Pham P, Top J, Willems RJL, van Schaik W, van Passel MWJ, Smidt H. 2017. Characterization of *Enterococcus* isolates colonizing the intestinal tract of intensive care unit patients receiving selective digestive decontamination. Front Microbiol 8:1596. https://doi.org/10.3389/fmicb.2017.01596.

49. Ghanem M, Wang L, Zhang Y, Edwards S, Lu A, Ley D, El-Gazzar M. 2018. Core genome multilocus sequence typing: a standardized approach for molecular typing of *Mycoplasma gallisepticum*. J Clin Microbiol 56:e01145-17. https://doi.org/10.1128/JCM.01145-17.

50. Bletz S, Janezic S, Harmsen D, Rupnik M, Mellmann A. 2018. Defining and evaluating a core genome multilocus sequence typing scheme for genome-wide typing of *Clostridium difficile*. J Clin Microbiol 56:e01987-17. https://doi.org/10.1128/JCM.01987-17.

51. Simon S, Trost E, Bender J, Fuchs S, Malorny B, Rabsch W, Prager R, Tietze E, Flieger A. 2018. Evaluation of WGS based approaches for investigating a food-borne outbreak caused by *Salmonella enterica* serovar Derby in Germany. Food Microbiol 71:46–54. https://doi.org/10.1016/j.fm.2017.08.017.

52. Leekitcharoenphon P, Nielsen EM, Kaas RS, Lund O, Aarestrup FM. 2014. Evaluation of whole genome sequencing for outbreak detection of *Salmonella enterica*. PLoS One 9:e87991. https://doi.org/10.1371/journal.pone.0087991.

53. Hammerum AM, Baig S, Kamel Y, Roer L, Pinholt M, Gumpert H, Holzknecht B, Røder B, Justesen US, Samulioniené J, Kjærsgaard M, Østergaard C, Holm A, Dzajic E, Søndergaard TS, Gaini S, Edquist P, Alm E, Lilje B, Westh H, Stegger M, Hasman H. 2017. Emergence of *vanA Enterococcus faecium* in Denmark, 2005–15. J Antimicrob Chemother 72:2184–2190. https://doi.org/10.1093/jac/dkx138.

54. Croucher NJ, Harris SR, Grad YH, Hanage WP. 2013. Bacterial genomes in epidemiology–present and future. Philos Trans R Soc Lond B Biol Sci 368:20120202. https://doi.org/10.1098/rstb.2012.0202.

55. Henri C, Leekitcharoenphon P, Carleton HA, Radomski N, Kaas RS, Mariet JF, Felten A, Aarestrup FM, Gerner Smidt P, Roussel S, Guillier L, Mistou MY, Hendriksen RS. 2017. An assessment of different genomic approaches for inferring phylogeny of *Listeria monocytogenes*. Front Microbiol 8:2351. https://doi.org/10.3389/fmicb.2017.02351.