



# Extensive Comparative Genomic Analysis of *Enterococcus faecalis* and *Enterococcus faecium* Reveals a Direct Association between the Absence of CRISPR–Cas Systems, the Presence of Anti-Endonuclease (*ardA*) and the Acquisition of Vancomycin Resistance in *E. faecium*

Kodjovi Mlaga, Vincent Garcia, Philippe Colson, Raymond Ruimy, Jean-Marc Rolain, Seydina Diene

## ► To cite this version:

Kodjovi Mlaga, Vincent Garcia, Philippe Colson, Raymond Ruimy, Jean-Marc Rolain, et al.. Extensive Comparative Genomic Analysis of *Enterococcus faecalis* and *Enterococcus faecium* Reveals a Direct Association between the Absence of CRISPR–Cas Systems, the Presence of Anti-Endonuclease (*ardA*) and the Acquisition of Vancomycin Resistance in *E. faecium*. *Microorganisms*, 2021, 9 (6), pp.1118. 10.3390/microorganisms9061118 . hal-03334105

**HAL Id: hal-03334105**

**<https://amu.hal.science/hal-03334105>**

Submitted on 9 Sep 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



## Article

# Extensive Comparative Genomic Analysis of *Enterococcus faecalis* and *Enterococcus faecium* Reveals a Direct Association between the Absence of CRISPR–Cas Systems, the Presence of Anti-Endonuclease (*ardA*) and the Acquisition of Vancomycin Resistance in *E. faecium*

Kodjovi D. Mlaga <sup>1</sup>, Vincent Garcia <sup>1</sup>, Philippe Colson <sup>1,2</sup>, Raymond Ruimy <sup>3</sup>, Jean-Marc Rolain <sup>1,2</sup> and Seydina M. Diene <sup>1,2,\*</sup>



**Citation:** Mlaga, K.D.; Garcia, V.; Colson, P.; Ruimy, R.; Rolain, J.-M.; Diene, S.M. Extensive Comparative Genomic Analysis of *Enterococcus faecalis* and *Enterococcus faecium* Reveals a Direct Association between the Absence of CRISPR–Cas Systems, the Presence of Anti-Endonuclease (*ardA*) and the Acquisition of Vancomycin Resistance in *E. faecium*. *Microorganisms* **2021**, *9*, 1118. <https://doi.org/10.3390/microorganisms9061118>

Academic Editors:  
Eleonora Giovanetti, Costas  
C. Papagiannitsis and Ibrahim Bitar

Received: 21 April 2021

Accepted: 19 May 2021

Published: 21 May 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

<sup>1</sup> Aix Marseille University, IRD, APHM, MEPHI, IHU-Mediterranee Infection, 13005 Marseille, France; mlagadodji@gmail.com (K.D.M.); vincent.garcia13@sfr.fr (V.G.); philippe.colson@univ-amu.fr (P.C.); jean-marc.rolain@univ-amu.fr (J.-M.R.)

<sup>2</sup> IHU-Mediterranee Infection, Aix-Marseille University, 13005 Marseille, France

<sup>3</sup> Department of Bacteriology at Nice Academic Hospital, Nice Medical University, 06003 Nice, France; ruimy.r@chu-nice.fr

\* Correspondence: seydina.diene@univ-amu.fr; Tel.: +33-(0)-4-91-32-43-75

**Abstract:** Here, we performed a comparative genomic analysis of all available genomes of *E. faecalis* ( $n = 1591$ ) and *E. faecium* ( $n = 1981$ ) and investigated the association between the presence or absence of CRISPR–Cas systems, endonuclease/anti-endonuclease systems and the acquisition of antimicrobial resistance, especially vancomycin resistance genes. Most of the analysed Enterococci were isolated from humans and less than 14% of them were from foods and animals. We analysed and detected CRISPR–Cas systems in 75.36% of *E. faecalis* genomes and only 4.89% of *E. faecium* genomes with a significant difference ( $p$ -value  $< 10^{-5}$ ). We found a negative correlation between the number of CRISPR–Cas systems and genome size ( $r = -0.397$ ,  $p$ -value  $< 10^{-5}$ ) and a positive correlation between the genome %GC content and the number of CRISPR–Cas systems ( $r = 0.215$ ,  $p$ -value  $< 10^{-5}$ ). Our findings showed that the presence of the anti-endonuclease *ardA* gene may explain the decrease in the number of CRISPR–Cas systems in *E. faecium*, known to deactivate the endonucleases' protective activities and enable the *E. faecium* genome to be versatile in acquiring mobile genetic elements, including carriers of antimicrobial resistance genes, especially *vanB*. Most importantly, we observed that there was a direct association between the absence of CRISPR–Cas, the presence of the anti-CRISPR *ardA* gene and the acquisition of vancomycin resistance genes.

**Keywords:** comparative genomics; *E. faecalis*; *E. faecium*; CRISPR–Cas; vancomycin resistance; recombinations

## 1. Introduction

Enterococci are an ancient genus of *Enterococcaceae* that have adapted to living in complex environments and surviving in harsh conditions [1,2]. The genus *Enterococcus* comprises 54 species that are ubiquitously present in the gastrointestinal tracts of animals, including mammals, reptiles, birds and insects, which are thought to be the largest reservoir of Enterococci [3]. Two species, *Enterococcus faecalis* and *Enterococcus faecium*, are the leading cause of the vast majority of hospital-acquired Enterococci infections in humans [4]. The plasticity of the *Enterococcus* genomes allows them to rapidly respond and adapt to the environment by acquiring genetic determinants. It increases their ability to colonise and infect their host and cause diseases [3]. The success of *E. faecium* and *E. faecalis* in evolving as multi-resistant nosocomial pathogens is associated with their capacity to harbour and spray adaptive genetic materials, including antimicrobial resistance genes encoded by

mobile genetic elements (MGEs) [5]. However, *E. faecium* is intrinsically more frequently reported as being more resistant to antibiotics—especially to vancomycin—than *E. faecalis* is (8.8% vs. 1.0% in Europe, 79.4% vs. 8.5% in the US and 22.4% vs. 0.1% in Canada) [6], and *vanA* and *vanB* are the most common mobile genes involved [7–10]. A sequence analysis of *E. faecalis* V583 revealed that 26% of the 3.36-Mb genome consisted of mobile elements, including 7 putative phages, 38 insertion elements and remnants of 3 integrated plasmids, as well as 3 independently replicating plasmids [4,11]. Enterococci genomic evolution has always been associated with the acquisition of vancomycin resistance genes carried by plasmids [12–17] and virulence genes [18–21]. The presence or absence of clustered, regularly interspaced, short palindromic repeats (CRISPR) can contribute to the regulation of this phenomenon and provide bacterial and archaeal immunity against subsequent invasion by plasmids and phages [22]. It is also known that restriction–modification systems and anti-endonuclease (*ardA*) play a significant role in the regulation of MGE transfer in the *Enterococcus* genus [11] and the acquisition and spread of antimicrobial resistance genes [23]. The aim of this study was to investigate the presence of recombination sites in both *E. faecalis* and *E. faecium* as evidence of MGE transfer and the association between the absence or presence of a CRISPR–Cas system, an endonuclease and anti-endonuclease system and the acquisition of antimicrobial resistance genes—especially the vancomycin resistance genes *vanA*, *vanB* and *vanC*—using sequenced genomes of *E. faecalis* from Marseille and publicly available genomes of both species.

## 2. Materials and Methods

### 2.1. Whole Genome Sequencing and Sequence Extraction from NCBI

We sequenced six genomes of *E. faecalis* composed of four strains isolated from humans and two from chicken faeces. The sequencing was initiated from the fact that an initial MALDI-TOF spectrum analysis revealed that they were closely related. We conducted whole genome nucleic acid extraction from the six strains using the QIAGEN automated method. We sequenced the *E. faecalis* genome using MiSeq Technology (Illumina Inc., San Diego, CA, USA) with the mate pair strategy [24]. The reads were assembled using A5-Miseq [25]. The scaffolds were re-ordered and aligned against a reference genome, *E. faecalis* ATCC 22809, using the Mauve aligner (v2.3.1) [26] with default parameters. For the purpose of the extensive comparative analysis, we extracted a total of 1591 whole genome sequences of *E. faecalis* and 1981 of *E. faecium* from the NCBI database. We used the `ncbi-genome-download` command with default parameters to download all available genomes (in fasta format) of *E. faecalis* and *E. faecium*, including their metadata. Only released genome sequences available for downloading were retrieved. Then, we thoroughly checked the metadata and excluded all the known contaminated sequences. Generally, the *E. faecalis* and *E. faecium* draft genome size is between 2.2 and 3.5 Mb (including plasmid sequences). In our analysis, whole genome sequences above 4 Mb with a percentage of aligned sequences (`mauve_Aligner`) < 50% were also excluded from our analysis. We re-annotated all 3572 genomes of the two species, including genomes sequenced from Marseille (strains G823, G824, G881, G882, G883 and G884 with assembly accession numbers FPDY01, FPDW01, FPEB01, FPDZ01, FPED01 and FPEA01, respectively), with Prokka [27] using Pfam and the Swissport database and default parameters.

### 2.2. Genome Phylogenetic Tree Reconstruction of Both *E. faecalis* and *E. faecium*

We performed a whole genome SNP alignment of the 1591 whole genomes of *E. faecalis* and the 1981 genomes of *E. faecium* using Scapper (<https://github.com/tseemann/scapper>, accessed on 10 January 2021), with default settings, to reconstruct the whole genome phylogenetic tree. The strains of the *E. faecalis* ATCC 29212 (NZ\_CP008816.1) and *E. faecium* DO (NC\_017960.1) genomes were used as references, respectively. Aligned multi-fasta were shrunk using SNP-sites [28] to remove the monomorphic sites, and so reduce the phylogenetic inference computing time. We inferred approximately maximum likelihood phylogenetic trees from shrunken alignments of nucleotide sequences using RaxML [29–31]

and used them to adjust the seed parameter to 2000 for a reproducible and consistent tree topology. The generated phylogenetic tree was used as the entry tree for ClonalFrameML analysis [32].

### 2.3. Detection of Recombination Hotspots inside the Genomes of *E. faecalis* and *E. faecium*

We putatively identified base substitution and recombination loci containing elevated densities of base substitutions suggestive of horizontal transferring sequences. We constructed a maximum likelihood phylogenetic tree based on the putative point mutations outside these regions of high sequence diversity using ClonalFrameML (version v1.0-19) [32]. We reconstructed the evolutionary maximum likelihood phylogeny by determining genetic genealogy and considering points of variation and genome plasticity and then generated a genome recombination heatmap against phylogenetic trees.

### 2.4. Detection of Clustered, Regularly Interspaced, Short Palindromic Repeat (CRISPR) Spacers inside the *E. faecalis* and *E. faecium* Genomes

Detection of a CRISPR–Cas system (spacers, repeats) was conducted using MinCED software (v0.4.1 [33]). We set the minimum number of repeats a CRISPR–Cas system must contain to three, the minimum length of the CRISPR repeats to 23 nucleotides, the maximum length of the CRISPR repeats to 47 nucleotides, the minimum length of the CRISPR spacers to 26 nucleotides and the maximum length of the CRISPR spacers to 50 nucleotides.

### 2.5. Orthologous Gene Detection and Pan-Genome Analysis of Both Species

The *pan-genome* analysis was performed using Roary [34] (version 3.6.8) with default parameters for orthologous gene similarity (BLAST: 95% minimum identity;  $E$ -value =  $1e^{-5}$ ), and genes had to be present in >99% of all isolates to be included in the hard-core genome. We determined core (hard-core) (genes present in 99–100% taxa), softcore (genes present in 95–99% taxa), shell (genes present in 15–95%) and cloud genes (genes present in 0–15% genomes) as described by Kaas [35]. We generated a binomial pan-genome profile for the presence, indicated as (1), and absence of genes, indicated as (0), inside the genome of both species. Orthologous genes related to antimicrobial, endonuclease and anti-endonuclease gene distribution were extracted using homemade scripts and plotted against maximum likelihood phylogenetic trees in both species.

### 2.6. Statistical Analysis

All statistical analyses conducted in this study were performed using R statistical software (v. 3.4.4) [36]. We used the Student's *t*-test for means' comparison, and the Pearson chi-squared test was used for proportion comparison inside and between the two species. The Pearson correlation test was used to show a statistical association between two genomic features. A logistic regression analysis was used to compute the association between qualitative genomic variables (presence and absence of vancomycin genes and the number of recombination and CRISPR spacers detected inside and across the genomes of both species). The odds ratio was calculated to interpret the association. We set the CI level to 95%. The statistical test was significant at a *p*-value of < 0.05. All *p*-values below 0.00001 were standardized as *p*-value <  $10^{-5}$  in this study.

## 3. Results

### 3.1. Comparison of the *E. faecalis* and *E. faecium* Genome Features Reveals Differences in Genome Size and No Difference in GC Percentage

We assembled the clinical Marseille strains G823, G824, G883 and G884 into 2.89, 3.096, 2.76 and 2.95 Mb, respectively, and the animal (chicken) strains G881 and G882 into 2.97 and 3.25 Mb, respectively. Overall, as presented on Figure S1A, 34.25% (545/1591) of *E. faecalis* genomes were from North America followed by 30.48% (485/1591) from Europe, and 34.42% (682/1981) of *E. faecium* genomes were from Europe. We noted that 31.99% (509/1591) of the *E. faecalis* strains and 55.12% (1092/1981) of the *E. faecium*

strains were isolated primarily from humans. However, a considerable number of *E. faecalis* (14.26%, 9.05%) and *E. faecium* (9.74%, 6.86%) were isolated from food and animals, respectively (Figure S1B). The average genome size of *E. faecalis* is estimated at 3.08 Mbp (min = 2.60 Mbp, max = 3.59 Mbp, sd = 0.15) and that of *E. faecium* at 2.91 Mbp (min = 2.23 Mbp, max = 3.72 Mbp, sd = 0.21), with a statistically significant difference between the two species ( $p$ -value <  $10^{-5}$ ). The average DNA coding sequence size of *E. faecalis* is estimated to be 2948 (min = 2357, max = 3423, sd = 173) and that of *E. faecium* to be 2708 (min = 2098, max = 4855, sd = 235). The GC percentage of *E. faecalis* is 37.42% (min = 36.80%, max = 39.67%, sd = 0.62), while that of *E. faecium* is 37.90% (min = 37.10%, max = 39.87%, sd = 0.43), with no significant difference ( $p$ -value = 3.97). The overall analysis of genome features indicates a statistically significant difference between *E. faecalis* and *E. faecium*. All available metadata associated to the analysed genomes are provided in Tables S1 and S2.

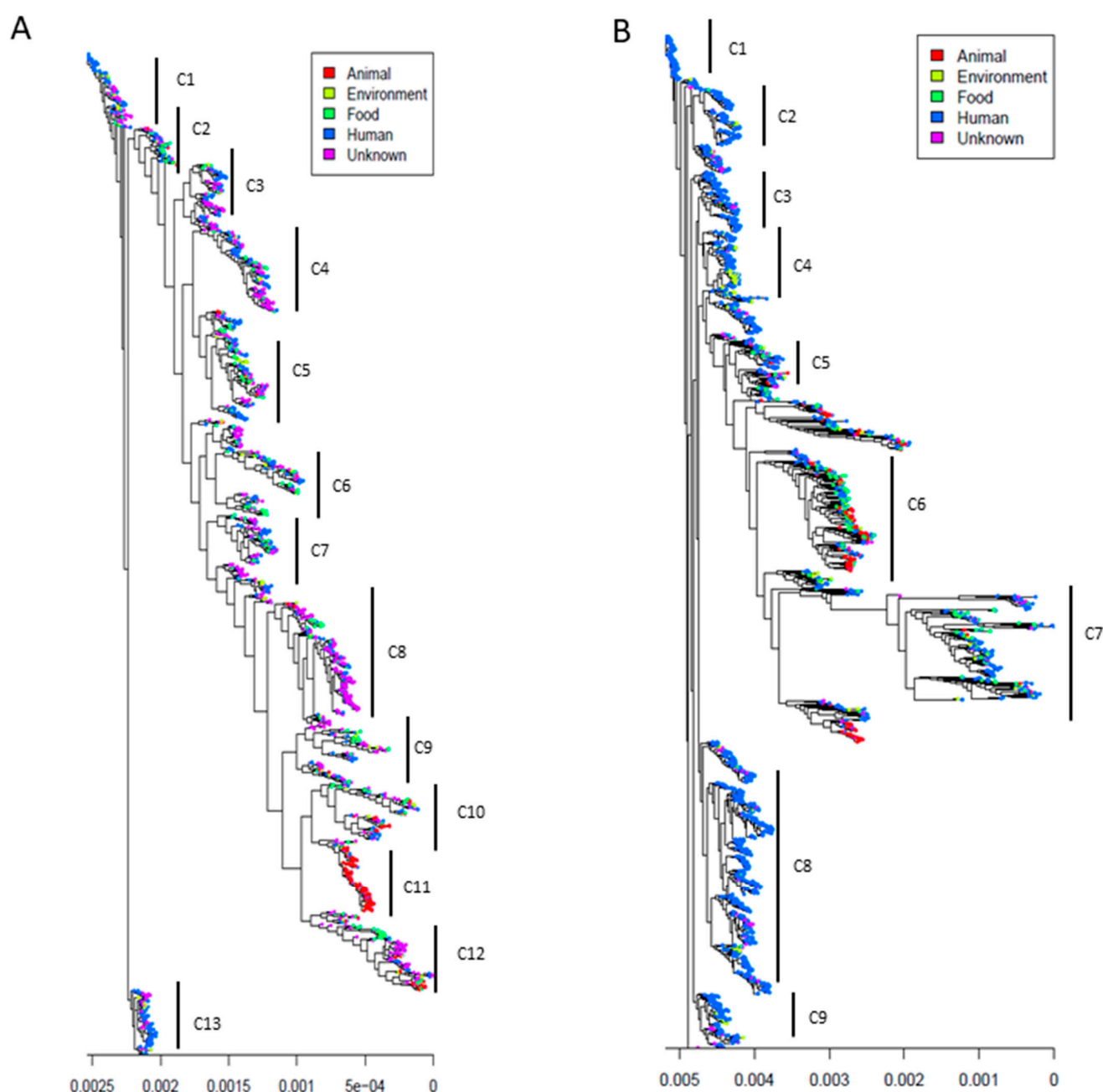
### 3.2. The Genomes of *E. faecalis* Contain a High Density of Recombination Hotspots Compared to *E. faecium*

We processed a whole genome SNP alignment for both species. The phylogenetic tree of *E. faecalis* shows 13 different small clades defined as a minimum of 20 to 30 branches sharing a common node (Figure 1A). With the exception of clade C11, which almost exclusively contains animal strains, and clade 13, which is predominantly composed of human strains, the remaining 11 clades consist of a mixture of human, animal, food and environmental strains. The length of the phylogenetic branches indicates the level of genomic evolution from the ancestral strain. Similar observations were made with *E. faecium*, with nine small heterogeneous clades (Figure 1B). As can be observed in Figure 1B, most of the *E. faecium* isolates were from humans.

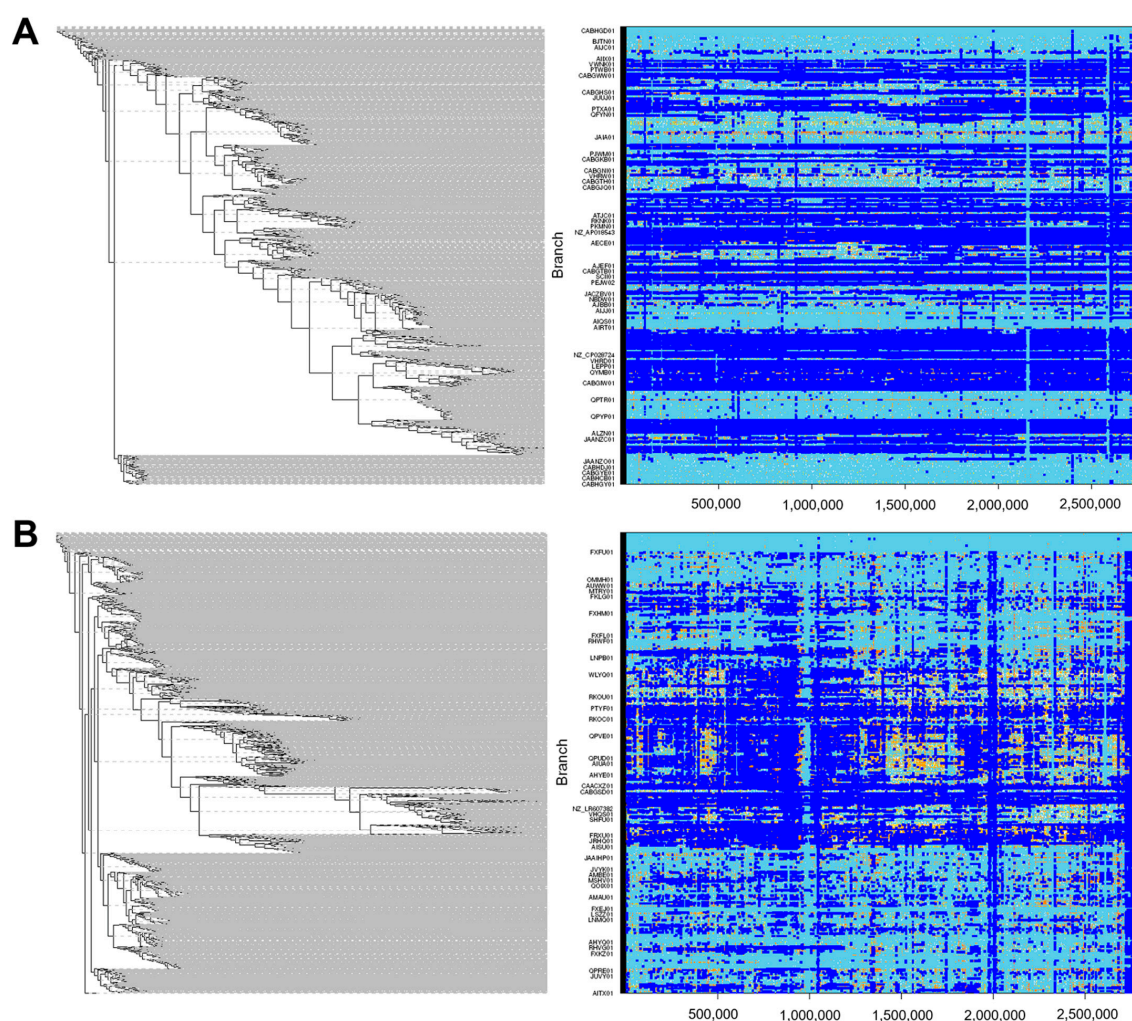
We detected recombination sites in 66% (1045/1591) of the genomes of *E. faecalis* that were analysed (Figure 2A). The number of recombination blocks is proportional to the length of the phylogenetic branch, which denotes a high level of recombination among these strains. Moreover, we observed that these strains contained a substantial number of animal strains. However, we detected recombination locks in only 30.94% (613/1981) of the *E. faecium* genomes (Figure 2B). We observed a similar profile as with *E. faecalis*, where most of the recombination occurred within the clades containing a mixture of isolates (animal, environmental, food and human). Overall, there were more genomic recombination hotspots detected in *E. faecalis* than in *E. faecium*, with a statistically significant difference ( $p$ -value <  $10^{-5}$ ). Moreover, most of these recombination hotspots occurred on branches with mixtures of strains from animals, food and the environment compared to those unique to humans.

As presented in Table S3, the pan-genome analysis identified a total of 39,665 orthologous genes in *E. faecalis* and 45,697 in *E. faecium*. Both species present a similar pan-genomic profile, with an exponential increase in the total number of orthologous genes and new genes when we added more genomes to the pan-genome. This suggests that the pan-genome of these two species is open. The hard-core genome of *E. faecalis* represents 30.73% of the average proteome, while that of *E. faecium* is estimated at 19.23%. A high number of orthologous genes were included in cloud genes (accessory genes)—36,161 and 41,885 for *E. faecalis* and *E. faecium*, respectively—suggesting a prominent level of genome plasticity.





**Figure 1.** Maximum likelihood phylogenetic tree of 1591 *E. faecalis* genomes and 1981 *E. faecium* genomes. (A): We detected 13 phylogenetic clades in *E. faecalis* (left); almost all clades are heterogenous, with a mixture of *E. faecalis* from humans, animals, foods, environment and unknown sources. Two clades, namely C11 and C13, appeared specific to animals and humans, respectively. Most of the *E. faecalis* strains are distant from ancestral strains. (B): In *E. faecium* (right), nine phylogenetic clades were detected. Despite the fact that all clades are composed of a combination of human and non-human strains, we can observe a predominance of *E. faecium* strains from human sources. Both phylogenetic trees were generated from an SNP alignment of the whole genome sequences. Monomorphic sites were trimmed from the alignment, and a first phylogenetic tree was produced using RAXML with the generalized reverse-time method. We used these trees as entering trees for ClonalFrameML analysis. Final midpoint-rooted clonal phylogenetic trees were generated.



**Figure 2.** Clonal phylogeny and inferred recombination events. Midpoint-rooted clonal phylogeny (left) of 1591 *E. faecalis* isolates (A) and 1981 *E. faecium* isolates (B). Recombination events (right) were estimated as described in the Method section. The sizes and genomic locations of recombination fragments (dark-blue line segments) occurring along branches in the phylogeny are aligned with branches in the phylogeny.

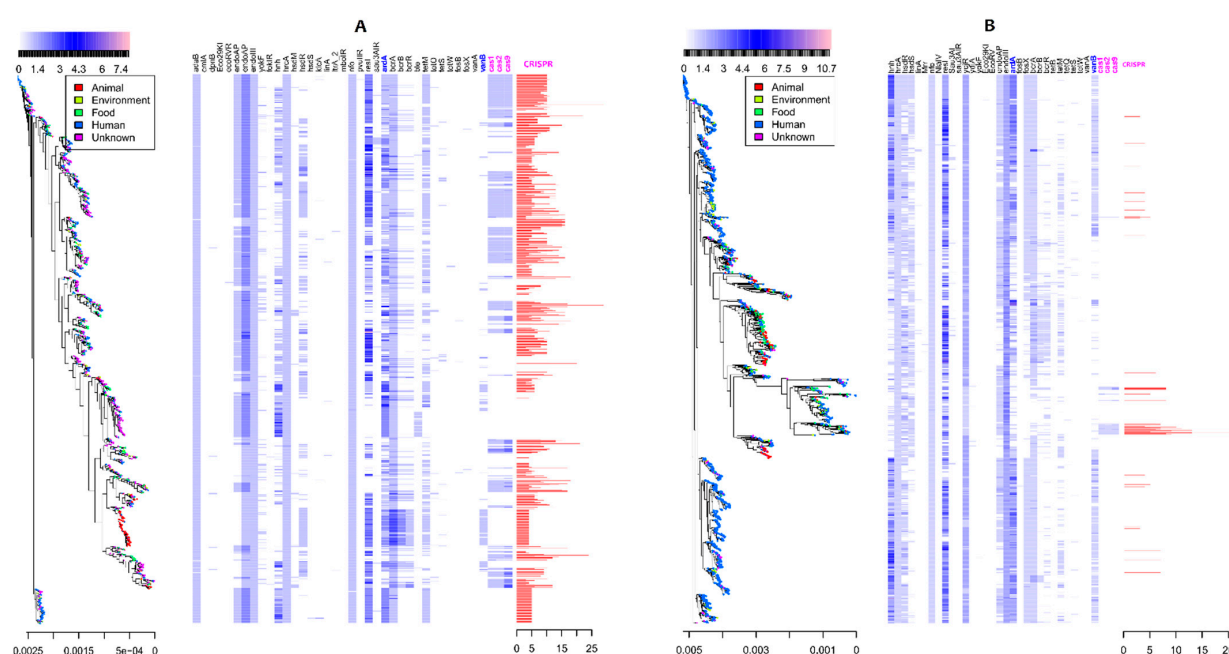
### 3.3. There Were More CRISPR–Cas Systems and Absence of Anti-Endonuclease in *E. faecalis* Genomes Compared to *E. faecium*

We detected CRISPR–Cas systems in 75.36% (1199/1591) of *E. faecalis* genomes and in only 4.89% (97/1981) of *E. faecium*, a statistically significant difference ( $p$ -value  $< 10^{-5}$ ). In *E. faecalis*, the number of CRISPR–Cas systems detected varied from 0 to 29 and contained five on average per genome. In *E. faecium*, it varied from 0 to 21, with less than one on average per genome. We found a positive association between the number of CRISPR spacers detected and the number of recombination blocks detected in *E. faecalis* (F-statistic: 14.39,  $p$ -value  $< 10^{-5}$ ). Three CRISPR–Cas-associated coding proteins—CRISPR–Cas1, CRISPR–Cas2 and CRISPR–Cas9—were identified in both species. We detected two variants of Cas9 (Cas9 and Cas9.1) in *E. faecalis* alone (sequence identity cut-off set at 80%).

### 3.4. The Association between the Absence of CRISPR–Cas Systems, the Presence of Anti-Endonuclease Genes (*ardA*) and Acquisition of the Vancomycin Resistance Genes *vanA*, *vanB* in *E. faecium*

We retrieved antimicrobial resistance genes, endonuclease genes, anti-endonuclease genes and presence and absence matrices from the overall pan-genome matrices, and the heatmaps were plotted against the maximum likelihood phylogeny trees to analyse their distribution in both *E. faecalis* (Figure 3A) and *E. faecium* (Figure 3B). We observed that

the vancomycin resistance gene *vanB* was present in both species. *E. faecium* harbours more vancomycin resistance genes than *E. faecalis* does (*E. faecium*/*E. faecalis*: 1069/201,  $p$ -value  $< 10^{-5}$ ). Most importantly, we detected the presence of vancomycin resistance genes in the genomes where a CRISPR–Cas system was absent in both species. Furthermore, endonuclease genes, including *cas1*, *cas2* and *cas9*, were found in both species, with a slight increase in *E. faecium*. However, anti-endonuclease genes (*ardA*) were found in large amounts in *E. faecium*. The presence of a CRISPR–Cas system in the genome of *E. faecium* decreased by 0.97 times the acquisition of vancomycin-resistant genes (estimate =  $-0.972$ , OR = 0.68,  $p$ -value  $< 10^{-5}$ , CI = [0.587–0.804]). The number of recombination hotspots detected in the genomes of both species decreased by 0.08 times the acquisition of vancomycin resistance genes (estimate =  $-0.08$ , OR = 0.98, CI = [0.956–0.995],  $p$ -value = 0.00021). Overall, there was a direct association between the absence of CRISPR spacers, the presence of anti-endonuclease genes (*ardA*) and the acquisition of vancomycin resistance in *E. faecium*.



**Figure 3.** Heatmap of antimicrobial orthologous genes detected in *E. faecalis* (A) and *E. faecium* (B) plotted against a maximum likelihood phylogenetic tree. Phylogenetic tree (left). Restriction enzymes/endonucleases and antimicrobial genes' distribution (centre), including CRISPR–Cas proteins; blue to pink ticks show the presence of orthologous genes in the corresponding genome on the phylogenetic tree; an empty space illustrates the absence of the genes. Red bars (right) show CRISPR spacers, with a scale indicating the presence and the number of spacers identified. Genes shown in the top horizontal bars are restriction enzymes/endonucleases/anti-endonucleases; 5-methylcytosine-specific restriction enzyme B (*mcrB*); anti-restriction protein (*ardA*); restriction endonuclease (*cfrBI*); restriction endonuclease (*eco29kI*); restriction-modification methylase (*eco57I*); EcoKI restriction-modification system protein (*hsdS*); 5-methylcytosine restriction system component (*mcrBC*); restriction system protein (*mrr*); restriction endonuclease (*ngoBV*); restriction endonuclease (*ngoFVII*); putative type-1 restriction enzyme specificity protein (*mg438*); putative type-1 restriction enzyme specificity protein (MPN\_089); restriction endonuclease (*sinI*); Type I restriction enzyme EcoKI M protein (*hsdM*); Type I restriction enzyme EcoKI M protein (*hsdM*); Type I restriction enzyme EcoKI M protein (*hsdR*); Type I restriction enzyme (*EcoR124II*); Type II restriction endonuclease (RE\_Alw26IDE); Type III restriction enzyme (Type III); Type II restriction enzyme (*sau3AI*); antimicrobial resistance genes: *bmrA* (multidrug resistance ABC transporter); *qacA* (multidrug efflux protein); *msbA* (lipid ABC transporter permease/ATPase; multidrug resistance ABC transporter); *ble* (bleomycin-resistant genes); *emrY* (multidrug resistance protein Y); *bcr* (bicyclomycin/multidrug efflux system); *yheH*, *yheI* (multidrug resistance ABC transporter); *fosB* (fosfomycin resistance gene B); *fosX* (fosfomycin resistance gene X); *tetA*, *tetC*, *tetD*, *tetM*, *tetR* (tetracycline resistance gene class A, C, D, M and R); *vanA* (vancomycin resistance gene A); *vanB* (vancomycin resistance genes B); *linA* (lincosamide B resistance genes); *cmlA* (chloramphenicol efflux protein); *cas1*, *cas2*, *cas9*, *cas9-1* (CRISPR-associated coding protein genes).



#### 4. Discussion

*E. faecalis* and *E. faecium* are the main causes of Enterococci nosocomial infections [3] and have been widely reported in blood [37] and in urinary tract infections [38]. In the past decade, the emergence of these two species has been attributed to resistance to antibiotics used in treating human infections, especially vancomycin [39]. Palmer et al. and Van Hal et al. demonstrated that vancomycin resistance genes, especially *vanA* and *vanB* and their associated regulatory genes *vanAHXZ* and *vanBYXZ*—are transferable by mobile genetic elements and plasmids [40,41]. Furthermore, Palmer et al. demonstrated that multidrug-resistant Enterococci lack the CRISPR–Cas protein in their genome [22]. Here, we performed an extensive comparative analysis of 3572 publicly available genomes of *E. faecalis* (1591) and *E. faecium* (1981), including six genomes of *E. faecalis* sequenced in Marseille. One limitation of our study can be identified regarding the fact that not all genomes were linked with their metadata; thus, there were a considerable number of genomes with missing metadata, such as the geographical origins and/or the isolation source of the strain. This includes 31.68% of the *E. faecalis* and 24% of the *E. faecium* genomes analysed in this study. Our study confirmed the high rate of vancomycin-resistant *E. faecium* reported around the world compared to *E. faecalis*, and the association with the presence or absence of a CRISPR–Cas system and its associated protein Cas in their respective genomes. The CRISPR–Cas system (spacers and related proteins) provides bacteria and archaea with a sequence-specific acquired defence system against plasmids and phage acquisition [42] and adaptive immunity against foreign elements. When the virus injects its genetic element into the bacteria, a small sequence of the viral genome, known as a spacer, is integrated into the CRISPR locus to immunize the bacteria. Spacers are transcribed into small RNAs that guide the direct cleavage of viral DNA by Cas nuclease proteins. The immunized population not only acquires resistance to its predators but also passes this resistance mechanism vertically to its progeny [42,43]. In this study, we observed that *E. faecalis* genomes contain a significantly higher number of recombination hotspots than *E. faecium* genomes do. *E. faecalis* is armed with a substantial number of CRISPR–Cas systems that protect the bacteria from acquiring subsequent external DNA, such as mobile genetic elements and plasmids. Therefore, we found a positive correlation between the number of recombination hotspots and the presence of CRISPR spacers found in *E. faecalis*. The massive *ardA* orthologues detected in *E. faecium* are evidence of the acquisition of plasmid elements, which are carriers of the vancomycin resistance genes *vanA* and *vanB* [10,39,44]. Our study also showed that there were more vancomycin resistance genes (*vanA* and/or *vanB*) detected in *E. faecium* than in *E. faecalis*, with a statistically significant difference. We found that the presence of a CRISPR–Cas system is protective for *E. faecalis* in acquiring specific mobile genetic elements carrying the vancomycin resistance genes *vanA* and *vanB*. However, the presence of *ardA* genes inactivates the function of endonuclease protective activities and enables the genome of *E. faecium* to be versatile in acquiring external DNA horizontally. The anti-endonuclease gene *ardA* is known to regulate horizontal gene transfer, causing multidrug resistance in *Enterococcus* [4,11] and actively contributing to the acquisition and dissemination of antimicrobial resistance genes [23]. These observations explain why *E. faecium* is often reported to exhibit more resistance to vancomycin than *E. faecalis* is. In this analysis, CRISPR–Cas9 was detected in *E. faecalis* as well as CRISPR–Cas1 and CRISPR–Cas2, and was almost absent in *E. faecium*. CRISPR–Cas1 and CRISPR–Cas2, two metal-dependent nucleases, are both necessary and sufficient for spacer acquisition, but dispensable for target interference [43,45]. However, CRISPR–Cas9, the sole Type II Cas protein, is involved in target surveillance and interference [46]. The endonuclease activity of Cas9 is dispensable for acquisition, as its role is to select spacers, whereas Cas1, whose non-specific nuclease activity is required for adaptation, cleaves the contiguous sequence, yielding a precisely selected spacer sequence [43]. The phylogenetic analysis revealed that some strains of *E. faecalis* and *E. faecium* isolated from humans share the same nodes with animal, food and environmental strains, suggesting that these strains may find their origin in animals and environments with zoonotic transmission. This

hypothesis is supported by Lebreton et al. [39], who discovered that an epidemic hospital-adapted lineage of *E. faecium* emerged approximately 75 years ago—concomitant with the introduction of antibiotics—from a population that included mostly animal strains, rather than from human commensal lines. The overuse or misuse of antimicrobial agents in animal farming can exert selection pressure on microbial communities and select resistant bacteria. These resistant bugs can pass from the environment and foods to humans through food and freshwater consumption. Humans develop infections due to multi-resistant bacteria that become difficult to treat in therapy. This explains why *E. faecium* strains are more associated with zoonotic dissemination, and the emergence of vancomycin-resistant *E. faecium* in humans may be related to the use of avoparcin as an animal growth promoter, known to produce cross-resistance to vancomycin [47].

## 5. Conclusions

This study shows that extensive genomic recombination has occurred in the *E. faecalis* species due to mobile genetic elements and phages capable of inducing adaptive immunity with the acquisition of a CRISPR–Cas system. This protects *E. faecalis* from acquiring external DNA sequences carrying the vancomycin resistance genes *vanA* and *vanB*. It correlates with the reduced number of CRISPR–Cas systems found in *E. faecium* and the substantial number of anti-endonuclease *ardA* genes and vancomycin resistance genes found. The emergence and dissemination of *E. faecium* infection may be due to zoonotic transmission, and the misuse of antibiotics (avoparcin) may cause the selection of emerging vancomycin resistance in Enterococci. This finding explains why *E. faecium* is more reported worldwide as a vancomycin-resistant *Enterococcus* species than *E. faecalis*.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/microorganisms9061118/s1>, Figure S1: Distribution of the geographical location and source of isolation of *E. faecalis* and *E. faecium*; Tables S1 and S2: List of all analysed genomes of *E. faecalis* and *E. faecium* with their metadata. Table S3: Distribution of pan-genome component in both *E. faecalis* ( $n = 1591$  genomes) and *E. faecium* ( $n = 1981$  genomes).

**Author Contributions:** Conceptualization, J.-M.R., R.R. and S.M.D.; methodology, K.D.M., S.M.D., V.G. and P.C.; formal analysis, K.D.M.; writing—original draft preparation, K.D.M.; writing—review and editing, K.D.M., S.M.D., P.C., R.R. and J.-M.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** The French government supported this work under the “Investissements d’avenir” programme (reference: Mediterranee Infection 10-IAHU-03).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Acknowledgments:** We are very grateful to the IHU Méditerranée-Infection, Marseille, France for financial support.

**Conflicts of Interest:** The authors report no conflict of interest or financial disclosures.

## References

1. Byappanahalli, M.N.; Nevers, M.B.; Korajkic, A.; Staley, Z.R.; Harwood, V.J. Enterococci in the environment. *Microbiol. Mol. Biol. Rev.* **2012**, *76*, 685–706. [CrossRef]
2. Elsner, H.A.; Sobottka, I.; Mack, D.; Claussen, M.; Laufs, R.; Wirth, R. Virulence factors of *Enterococcus faecalis* and *Enterococcus faecium* blood culture isolates. *Eur. J. Clin. Microbiol. Infect. Dis.* **2000**, *19*, 39–42. [CrossRef] [PubMed]
3. Guzman Prieto, A.M.; van Schaik, W.; Rogers, M.R.C.; Coque, T.M.; Baquero, F.; Corander, J.; Willems, R.J.L. Global Emergence and Dissemination of Enterococci as Nosocomial Pathogens: Attack of the Clones? *Front. Microbiol.* **2016**, *7*, 1–15. [CrossRef] [PubMed]
4. Higueta, N.I.A.; Huycke, M.M. Enterococcal Disease, Epidemiology, and Implications for Treatment. In *Enterococci: From Commensals to Leading Causes of Drug Resistant Infection*; Gilmore, M.S., Clewell, D.B., Ike, Y.S.N., Eds.; Massachusetts Eye and Ear Infirmary: Boston, MA, USA, 2014; pp. 1–27.

5. García-Solache, M.; Lebreton, F.; McLaughlin, R.E.; Whiteaker, J.D.; Gilmore, M.S.; Rice, L.B. Homologous Recombination within Large Chromosomal Regions Facilitates Acquisition of  $\beta$ -Lactam and Vancomycin Resistance in *Enterococcus faecium*. *Antimicrob. Agents Chemother.* **2016**, *60*, 5777–5786. [[CrossRef](#)]
6. Kristich, C.J.; Rice, L.B.; Arias, C.A. *Enterococcal Infection—Treatment and Antibiotic Resistance*; Massachusetts Eye and Ear Infirmary: Boston, MA, USA, 2014.
7. Bourgogne, A.; Garsin, D.A.; Qin, X.; Singh, K.V.; Sillanpaa, J.; Yerrapragada, S.; Ding, Y.; Dugan-Rocha, S.; Buhay, C.; Shen, H.; et al. Large scale variation in *Enterococcus faecalis* illustrated by the genome analysis of strain OG1RF. *Genome Biol.* **2008**, *9*, R110. [[CrossRef](#)] [[PubMed](#)]
8. Solheim, M.; Brekke, M.C.; Snipen, L.G.; Willems, R.J.L.; Nes, I.F.; Brede, D. A Comparative genomic analysis reveals significant enrichment of mobile genetic elements and genes encoding surface structure-proteins in hospital-associated clonal complex 2 *Enterococcus faecalis*. *BMC Microbiol.* **2011**, *11*, 3. [[CrossRef](#)] [[PubMed](#)]
9. Do Prado, G.V.B.; Marchi, A.P.; Moreno, L.Z.; Rizek, C.; Amigo, U.; Moreno, A.M.; Rossi, F.; Guimaraes, T.; Levin, A.S.; Costa, S.F. Virulence and resistance pattern of a novel sequence type of linezolid-resistant *Enterococcus faecium* identified by whole-genome sequencing. *J. Glob. Antimicrob. Resist.* **2016**, *6*, 27–31. [[CrossRef](#)]
10. Sivertsen, A.; Pedersen, T.; Larssen, K.W.; Bergh, K.; Rønning, T.G.; Radtke, A.; Hegstad, K. A Silenced *vanA* Gene Cluster on a Transferable Plasmid Caused an Outbreak of Vancomycin-Variable Enterococci. *Antimicrob. Agents Chemother.* **2016**, *60*, 4119–4127. [[CrossRef](#)]
11. Palmer, K.L.; van Schaik, W.; Willems, R.J.L.; Gilmore, M.S.; Willem Van Schaik, C.A.; Willems, R.J.L.; Gilmore, M.S. *Enterococcal Genomics*; Massachusetts Eye and Ear Infirmary: Boston, MA, USA, 2014.
12. Evans, M.; Davies, J.K.; Sundqvist, G.; Figdor, D. Mechanisms involved in the resistance of *Enterococcus faecalis* to calcium hydroxide. *Aust. Endod. J.* **2001**, *27*, 115. [[CrossRef](#)]
13. Paulsen, I.T.; Banerjee, L.; Myers, G.S.A.; Nelson, K.E.; Seshadri, R.; Read, T.D.; Eisen, J.A.; Gill, S.R.; Fouts, D.E.; Heidelberg, J.F. Role of mobile DNA in the evolution of vancomycin-resistant *Enterococcus faecalis*. *Science* **2003**, *299*, 2071–2074. [[CrossRef](#)]
14. Hegstad, K.; Mikalsen, T.; Coque, T.M.; Werner, G.; Sundsfjord, A. Mobile genetic elements and their contribution to the emergence of antimicrobial resistant *Enterococcus faecalis* and *Enterococcus faecium*. *Clin. Microbiol. Infect.* **2010**, *16*, 541–554. [[CrossRef](#)]
15. Lebreton, F.; Depardieu, F.; Bourdon, N.; Fines-Guyon, M.; Berger, P.; Camiade, S.; Leclercq, R.; Courvalin, P.; Cattoir, V. D-Ala-D-Ser VanN-Type Transferable Vancomycin Resistance in *Enterococcus faecium*. *Antimicrob. Agents Chemother.* **2011**, *55*, 4606–4612. [[CrossRef](#)] [[PubMed](#)]
16. Arias, C.A.; Murray, B.E. The rise of the Enterococcus: Beyond vancomycin resistance. *Nat. Rev. Microbiol.* **2012**, *10*, 266–278. [[CrossRef](#)]
17. Mikalsen, T.; Pedersen, T.; Willems, R.; Coque, T.M.; Werner, G.; Sadowy, E.; van Schaik, W.; Jensen, L.B.; Sundsfjord, A.; Hegstad, K. Investigating the mobilome in clinically important lineages of *Enterococcus faecium* and *Enterococcus faecalis*. *BMC Genom.* **2015**, *16*, 282. [[CrossRef](#)]
18. Kayaoglu, G.; Ørstavik, D. Virulence factors of *Enterococcus faecalis*: Relationship to endodontic disease. *Crit. Rev. Oral Biol. Med.* **2004**, *15*, 308–320. [[CrossRef](#)]
19. Nallapareddy, S.R.; Singh, K.V.; Sillanp, J.; Garsin, D.A.; Hk, M.; Erlandsen, S.L.; Murray, B.E. Endocarditis and biofilm-associated pili of *Enterococcus faecalis*. *J. Clin. Investig.* **2006**, *116*, 2799–2807. [[CrossRef](#)] [[PubMed](#)]
20. Domann, E.; Hain, T.; Ghai, R.; Billion, A.; Kuenne, C.; Zimmermann, K.; Chakraborty, T. Comparative genomic analysis for the presence of potential enterococcal virulence factors in the probiotic *Enterococcus faecalis* strain Symbioflor 1. *Int. J. Med. Microbiol.* **2007**, *297*, 533–539. [[CrossRef](#)]
21. Soheili, S.; Ghafourian, S.; Sekawi, Z.; Neela, V.; Sadeghifard, N.; Ramli, R.; Hamat, R.A. Wide distribution of virulence genes among *Enterococcus faecium* and *Enterococcus faecalis* clinical isolates. *Sci. World J.* **2014**, *2014*, 623174. [[CrossRef](#)] [[PubMed](#)]
22. Palmer, K.L.; Kos, V.N.; Gilmore, M.S. Horizontal gene transfer and the genomics of enterococcal antibiotic resistance. *Curr. Opin. Microbiol.* **2010**, *13*, 632–639. [[CrossRef](#)]
23. McMahon, S.A.; Roberts, G.A.; Johnson, K.A.; Cooper, L.P.; Liu, H.; White, J.H.; Carter, L.G.; Sanghvi, B.; Oke, M.; Walkinshaw, M.D.; et al. Extensive DNA mimicry by the ArdA anti-restriction protein and its role in the spread of antibiotic resistance. *Nucleic Acids Res.* **2009**, *37*, 4887–4897. [[CrossRef](#)]
24. Ravi, R.K.; Walton, K.; Khosroheidari, M. MiSeq: A Next Generation Sequencing Platform for Genomic Analysis. *Methods Mol. Biol.* **2018**, 223–232. [[CrossRef](#)]
25. Coil, D.; Jospin, G.; Darling, A.E. A5-miseq: An updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics* **2015**, *31*, 587–589. [[CrossRef](#)]
26. Darling, A.C.E.; Mau, B.; Blattner, F.R.; Perna, N.T. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **2004**, *14*, 1394–1403. [[CrossRef](#)]
27. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **2014**, *30*, 2068–2069. [[CrossRef](#)] [[PubMed](#)]
28. Keane, J.A.; Page, A.J.; Delaney, A.J.; Taylor, B.; Seemann, T.; Harris, S.R.; Soares, J. SNP-sites: Rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb. Genom.* **2016**, *2*. [[CrossRef](#)]
29. Stamatakis, A. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **2006**, *22*, 2688–2690. [[CrossRef](#)]

- 
30. Minh, B.Q.; Nguyen, M.A.T.; von Haeseler, A.; Quang Minh, B.; Anh Thi Nguyen, M.; von Haeseler, A. Ultrafast Approximation for Phylogenetic Bootstrap. *Mol. Biol. Evol.* **2013**, *30*, 1188–1195. [[CrossRef](#)]
  31. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*. [[CrossRef](#)] [[PubMed](#)]
  32. Didelot, X.; Wilson, D.J. ClonalFrameML: Efficient inference of recombination in whole bacterial genomes. *PLoS Comput. Biol.* **2015**, *11*, e1004041. [[CrossRef](#)]
  33. Minced: Mining CRISPRs in Environmental Datasets. Available online: <https://github.com/ctSkennerton/minced/tree/master> (accessed on 10 January 2021).
  34. Page, A.J.; Cummins, C.A.; Hunt, M.; Wong, V.K.; Reuter, S.; Holden, M.T.G.; Fookes, M.; Falush, D.; Keane, J.A.; Parkhill, J. Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **2015**, *31*. [[CrossRef](#)]
  35. Kaas, R.S.; Friis, C.; Ussery, D.W.; Aarestrup, F.M.; Otto, T.; Oryan, M.; Prado, V.; Pickering, L.; Russo, T.; Johnson, J.; et al. Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse *Escherichia coli* genomes. *BMC Genom.* **2012**, *13*, 577. [[CrossRef](#)]
  36. R Development Core Team R: A language and environment for statistical computing. *Stat. Comput.* **2006**. [[CrossRef](#)]
  37. Furtado, I.; Xavier, P.C.N.; Tavares, L.V.M.; Alves, F.; Martins, S.F.; de Martins, A.S.; Palhares, D.B. *Enterococcus faecium* and *Enterococcus faecalis* in blood of newborns with suspected nosocomial infection. *Rev. Inst. Med. Trop. São Paulo* **2014**, *56*, 77–80. [[CrossRef](#)]
  38. Ronald, A. The etiology of urinary tract infection: Traditional and emerging pathogens. *Dis. Mon.* **2003**, *49*, 71–82. [[CrossRef](#)]
  39. Lebreton, F.; van Schaik, W.; McGuire, A.M.; Godfrey, P.; Griggs, A.; Mazumdar, V.; Corander, J.; Cheng, L.; Saif, S.; Young, S.; et al. Emergence of epidemic multidrug-resistant *Enterococcus faecium* from animal and commensal strains. *MBio* **2013**, *4*. [[CrossRef](#)] [[PubMed](#)]
  40. Palmer, K.L.; Carniol, K.; Manson, J.M.; Heiman, D.; Shea, T.; Young, S.; Zeng, Q.; Gevers, D.; Feldgarden, M.; Birren, B.; et al. High-quality draft genome sequences of 28 *Enterococcus* sp. isolates. *J. Bacteriol.* **2010**, *192*. [[CrossRef](#)] [[PubMed](#)]
  41. Van Hal, S.J.; Ip, C.L.C.; Ansari, M.A.; Wilson, D.J.; Espedido, B.A.; Jensen, S.O.; Bowden, R. Evolutionary dynamics of *Enterococcus faecium* reveals complex genomic relationships between isolates with independent emergence of vancomycin resistance. *Microb. Genom.* **2016**, *2*. [[CrossRef](#)]
  42. Marraffini, L.A. CRISPR-Cas immunity in prokaryotes. *Nature* **2015**, *526*, 55–61. [[CrossRef](#)] [[PubMed](#)]
  43. Ratner, H.K.; Sampson, T.R.; Weiss, D.S. Overview of CRISPR-Cas9 Biology. *Cold Spring Harb. Protoc.* **2016**, *2016*. [[CrossRef](#)]
  44. Brodrick, H.J.; Raven, K.E.; Harrison, E.M.; Blane, B.; Reuter, S.; Török, M.E.; Parkhill, J.; Peacock, S.J. Whole-genome sequencing reveals transmission of vancomycin-resistant *Enterococcus faecium* in a healthcare network. *Genome Med.* **2016**, *8*, 4. [[CrossRef](#)]
  45. Datsenko, K.A.; Pougach, K.; Tikhonov, A.; Wanner, B.L.; Severinov, K.; Semenova, E. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat. Commun.* **2012**, *3*, 945. [[CrossRef](#)] [[PubMed](#)]
  46. Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J.A.; Charpentier, E. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **2012**, *337*. [[CrossRef](#)] [[PubMed](#)]
  47. Nilsson, O. Vancomycin resistant enterococci in farm animals—Occurrence and importance. *Infect. Ecol. Epidemiol.* **2012**, *2*. [[CrossRef](#)] [[PubMed](#)]