applied sciences

MDPI

Article

Whole-Genome Sequencing and Comparative Genomic Analysis of *Enterococcus* spp. Isolated from Dairy Products: Genomic Diversity, Functional Characteristics, and Pathogenic Potential

Ilias Apostolakos [1], Markella Tsigkrimani [2], Spiros Paramithiotis [2] and Marios Mataragas [1,*]

[1] Department of Dairy Research, Institute of Technology of Agricultural Products, Hellenic Agricultural Organization "DIMITRA", 3 Ethnikis Antistaseos St., 45221 Ioannina, Greece
[2] Department of Food Science and Human Nutrition, Agricultural University of Athens, 75 Iera Odos St., 11855 Athens, Greece
* Correspondence: mmatster@elgo.gr; Tel.: +30-2651094780

Abstract: Enterococci are commensal organisms that have probiotic effects for their hosts and can be used as adjunct cultures in fermented dairy products. The dark side of Enterococci is manifested in *E. faecium* and *E. faecalis*, which are the causative agents of nosocomial infections, and thus Enterococci have not been granted a safety status as food additives. In this context, we aimed to assess the safety and functional profile of an Enterococci collection retrieved from traditional dairy products through a high-resolution genomic characterization and comparative genomic analysis. Analysis did not reveal major differences in the main cellular processes of Enterococci. Moreover, a diverse repertoire of resistance and virulence genes was present, though known hallmark pathogenicity factors were either absent or occurred rarely. The abundance of bacteriocins and CRISPR/Cas systems suggested the ability of the isolates to suppress pathogens and evade bacteriophages, respectively. Presence–absence patterns of genes suggested that dairy-originated *E. faecium* are not associated with pathogenicity factors, while those of human origin are strongly linked with notorious resistance and virulence determinants. Our comparative analysis provided some notable insights regarding the genomic composition of Enterococci in the context of their origin. However, their pathogenic lifestyle is likely to be explained by the interplay of multiple genomic factors.

Keywords: *E. faecalis*; *E. faecium*; next generation sequencing; safety assessment; resistance; virulence

Citation: Apostolakos, I.; Tsigkrimani, M.; Paramithiotis, S.; Mataragas, M. Whole-Genome Sequencing and Comparative Genomic Analysis of *Enterococcus* spp. Isolated from Dairy Products: Genomic Diversity, Functional Characteristics, and Pathogenic Potential. *Appl. Sci.* **2022**, *12*, 9620. https://doi.org/10.3390/app12199620

Academic Editor: Wojciech Kolanowski

Received: 30 August 2022
Accepted: 23 September 2022
Published: 25 September 2022

1. Introduction

Enterococci are Gram-positive, facultative, anaerobic bacteria. They are commensal inhabitants of the intestinal tract of humans and animals. However, some species of the genus *Enterococcus*, such as *E. faecalis* and *E. faecium*, can be opportunistic pathogens causing various diseases in humans and animals. In humans, Enterococci are mainly implicated in nosocomial bloodstream and urinary tract infections (UTIs) [1]. In food-producing animals, diseases caused by *Enterococcus* spp. are rare, although they are occasionally associated with mastitis in dairy animals [2]. Coagulase negative Staphylococci, and Streptococci, are the predominant causative agents of contagious mastitis, whereas Enterococci are one of the environmental pathogens, collectively implicated in 6% and 4% of mastitis cases in dairy ewes and goats, respectively [3].

Enterococci have inherent mechanisms of antimicrobial resistance (AMR) to several antimicrobial drugs. Moreover, they can acquire resistance mechanisms by spontaneous mutations or though horizontal gene transfer (i.e., exchange of mobile genetic elements between bacteria, such as plasmids) [4]. Ampicillin, aminoglycosides, and vancomycin resistance is common among *E. faecium* and *E. faecalis* isolates responsible for nosocomial

infections. The number of infections caused by vancomycin-resistant Enterococci has risen dramatically in recent years, which is alarming considering that vancomycin is a last-resort antimicrobial against a wide spectrum of multidrug-resistant (MDR), Gram-positive bacteria [5]. Studies suggest that food-producing animals are probably not a significant reservoir of human disease given that the AMR patterns in animal isolates are different from those isolated from human clinical cases [6]. Nevertheless, there are increasing concerns that important mechanisms conferring resistance to critical antimicrobials in human medicine can be disseminated to other bacteria that populate the gastrointestinal tract of animals. These bacteria, including those that can cause human disease, can reach consumers through the food chain [7].

As with many bacteria, there are two sides to Enterococci. Several strains are considered beneficial for their hosts by exerting their probiotics properties in vivo, such as strain T110 of *E. faecium*, while others have been used as adjunct cultures in fermenting dairy products [8,9]. Previous studies have shown that naturally occurring *Enterococcus* spp. in dairy products significantly contribute to the complex metabolic network created during the ripening process that eventually affects the organoleptic properties of the end products. Furthermore, they exert antimicrobial activity through the production of bacteriocins and thus have the capacity to modulate the microbiota of dairy products and control pathogens [7]. Despite the potential beneficial effects, none of the *Enterococcus* species has been granted the Qualified Presumption of Safety (QPS) status by the European Food Safety Agency (EFSA) [9] or the GRAS (Generally Regarded as Safe) USA equivalent. Therefore, there is a need to employ high-throughput technologies that can help to distinguish friend from foe in Enterococci via their thorough characterization. In this regard, the aim of this study was to provide a high-resolution characterization of *Enterococcus* spp. isolated from different dairy products [10,11], such as raw sheep milk, artisanal Feta, and artisanal Kefalograviera cheese, using whole-genome sequencing (WGS), primarily with respect to their resistance and virulence repertoire, but considering other important genomic features as well. A comparative genomic analysis of *Enterococcus* spp. isolated from various sources, including human clinical isolates, was conducted as well to assess the genomic characteristics of this collection in the context of a broader and diverse set of sequenced isolates.

2. Materials and Methods

2.1. Microbial Strains and Culture Conditions

The *Enterococcus* microbial collection (*E. faecium* = 24, *E. faecalis* = 3, *E. italicus* = 2, and *E. durans* = 1) of the dairy research department (DRD) of Hellenic Agricultural Organization "DIMITRA" (ELGO-DIMITRA), isolated from sheep milk and artisanal Feta and Kefalograviera cheeses [10], was used in this work. Storage and culture conditions of the strains are described in detail in the work of Tsigkrimani et al. [10,11].

2.2. Whole Genome Sequencing, Assembly, and Quality Control

DNA extraction was performed according to Syrokou et al. [12] and DNA sequencing by Novogene Genomics Service (Novogene Co., Cambridge, UK). Degradation and contamination of DNA was examined by agarose gel electrophoresis while quantification by Qubit 2.0 (ThermoFisher Scientific, Waltham, MA, USA). Library construction was performed as follows: sonication was employed for random fragmentation of DNA, followed by end polishing, A-tailing, ligation with the Illumina sequencing adapters, and finally PCR amplification with P5 and P7 oligos. Purification of the PCR products was performed with the AMPure XP system (Beckman Coulter, Brea, CA, USA), size distribution assessment with the Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA), and quantification by qPCR. Sequencing of the qualified libraries took place in the Illumina Novaseq 6000 sequencer (Illumina, San Diego, CA, USA) (2 × 150 bp). FastQC v.0.11 tool of the KBase platform [13,14] was employed for quality assessment of the adapter-free raw reads. Polishing and de novo assembling of the reads into contigs took place using the Pilon tool and the Unicycler assem-

bler, respectively, provided by the PATRIC v3.6.8 web platform [15–17]. The Multi-Draft based Scaffolder (MeDuSa) v1.6 platform [18] was used to assemble the contigs into scaffolds. The latter were then ordered and oriented on the basis of multiple complete reference genomes present in the NCBI database (https://www.ncbi.nlm.nih.gov/ accessed on 10 January 2022); *E. faecium* ATCC8459, *E. faecium* NBRC100486, *E. faecium* SRR24, *E. faecium* LMG8148, *E. faecium* NCTC7171, and *E. faecium* NRRLB-2354 (for *E. faecium*); *E. faecalis* ATCC19433, *E. faecalis* EnGen0336_T5, *E. faecalis* NCTC775, and *E. faecalis* NBRC100480 (for *E. faecalis*); *E. durans* NBRC100479, *E. durans* ATCC6056, *E. durans* NCTC8130; *E. durans* NCTC8129, and *E. durans* 8L1-82 (for *E. durans*); and *E. italicus* DSM15952 (for *E. italicus*). The CheckM tool v1.21 [19] of the PATRIC v3.6.8. platform was used for quality assessment of the contigs and scaffolds to ensure that assembled genomes ere of high quality, i.e., completeness ($\geq$95%) and contamination ($\leq$5%). The Skew Index Test (SkweIT) v1.0 web tool [20] was used to assess possible mis-assemblies after scaffolding.

2.3. In Silico Typing and Characterization

Genome assembly statistics were assessed with QUAST v5.2.0 [21]. Species identification was performed with the Kraken2 v2.1.2 taxonomic classifier [22] and the Type Strain Genome Server (TYGS) [23]. The genomes were annotated using PROKKA v1.14.5 [24] while further functional annotation and subsystem analysis of predicted open reading frames (ORFs) was done via the COG database (March 2022 version) [25]. Moreover, the presence of clustered regularly interspaced short palindromic repeats (CRISPRs) was evaluated with CRISPRCas-Finder v4.2.20 [26], while integrated prophages were identified with PHASTER (December 2020 version) [27]. Abricate v1.0.0 [28] was used to determine the presence of resistance genes (RGs), virulence genes (VGs) and mobile genetic elements (MGEs) using the Resfinder (May 2022 version) [29], VFDB (August 2022 version) [30] and MobileElementFinder (June 2020 version) [31] databases, respectively. Additionally, the presence of bacteriocins was determined with BAGEL4 v2.0-115 [32], whereas plasmids were detected, reconstructed, and genotyped with MOB-suite v.3.1.0 [33]. Lastly, we used the PathogenFinder v.1.1 [34] classifier to predict the pathogenicity of the isolates in our collection.

2.4. Phylogenetic Analysis and Comparative Genomics

Multilocus sequence types (MLSTs) of *E. faecium* isolates were identified via the PubMLST database (https://pubmlst.org/ accessed on 11 May 2022). In addition to the 24 *E. faecium* isolates of this study, 161 isolates from various sources were downloaded from the NCBI database to conduct a comparative genomic analysis. This selection was created by the investigation of all high quality (excluding "anomalous" filter) and taxonomically accurate (taxonomy "OK" filter) *E. faecium* genomes available at the NCBI assembly database. Only bacterial genomes with robust documentation of their isolation source and isolated from human, animal or food sources were included. The pangenome analysis and core-genome alignment of all *E. faecium* isolates was performed with Roary [35]. Proteins were assigned to the same family if their amino acid sequence identity was $\geq$95%. The threshold percentage of the isolates that needed to have a gene in order this to be considered as a core gene was set at 95%. Moreover, regions indicative of homologous recombination were removed with Gubbins [36] and a phylogenetic tree was built with FastTree [37]. Cluster calculation was performed with the hierBAPS module of the Bayesian Analysis of Population Structure (BAPS) software v6.0 [38]. BAPS clusters were assigned based on the core-genome alignment with 2 levels of hierarchy. The phylogenetic tree was annotated and visualized using the Interactive Tree Of Life (iTOL) program [39].

Furthermore, we conducted Carbohydrate-active enzyme (CAZyme) searches with the Run_dbcan V3 standalone tool of the dbCAN2 server [40], considering as positive hits only the genes found by both the Pfam Hidden Markov Models (HMMs) and DIAMOND. To further elucidate key genomic differences between *E. faecium* of human and dairy origin, a cluster heatmap was generated using a presence-absence matrix of the RGs, VGs and CAZymes present in these isolates. Clustering observations on the heatmap were further

explored with statistical analysis for gene enrichment of these gene classes in the respective isolate sources. Lastly, isolates of human and dairy origin were juxtaposed based on various predicted phenotypic traits (n = 67), using Traitar [41].

2.5. Statistical Analysis

One-way ANOVA and Tukey's post-hoc test implemented in R programming language [42] was used to determine statistically significant differences for the average genome size, GC content and number of coding sequences (CDSs) between *Enterococcus* species and difference in the total number of RGs between *E. faecium* origins. Statistical tests for significant genomic differences were conducted between species or groups having at least two isolates. Moreover, we used presence-absence data matrix of the RGs, VGs and CAZymes as input to Scoary [43] to analyze their enrichment in each isolation source. The significance level (alpha) was set at 0.05. The *p*-values were adjusted with the Benjamini-Hochberg method for multiple comparisons correction. Cluster heatmaps were generated in R using the pheatmap package.

3. Results

3.1. Species Identification, Assembly Statistics and Subsystem Analysis

Taxonomic classification with Kraken2 and TYGS corroborated the strain identification of Tsigkrimani, et al. [10,11]. From the 30 strains included in this analysis, 25 (83.3%) were classified as *E. faecium*, while three isolates of *E. faecalis*, two isolates of *E. italicus*, and one isolate of *E. durans* were found. Details and assembly statistics are presented in Table 1. The average genome size was 2.79 ± 0.15 million bases (Mb), the average GC content was 37.92% ± 0.40%, while the average number of coding sequences (CDSs) was 2850 ± 179, where ± denotes the standard deviation of the average. No significant difference between *Enterococcus* species was found for these genomic features. Subsystem analysis with the COG database revealed the presence of 11 enriched subsystem categories (Figure 1). The process category of metabolism was the most enriched one with 333 (±40) genes, on average. This category was significantly more enriched ($p < 0.05$) in the isolates of *E. faecalis* compared with those of *E. faecium* (ent_C28, ent_C78 and ent_C158). Together with metabolism, protein processing (209 ± 5), energy (120 ± 10), and stress-response-virulence (101 ± 5) processes constituted the majority of CDSs with known functions (Figure 1).

Table 1. Species identification and assembly statistics of the 30 *Enterococcus* spp. isolates.

Strain ID	Genus & Species	Genome Size (Mb)	GC Content (%)	No of Scaffolds	N50 (Mb)	No of CDSs	Contamination (%)	Completeness (%)
ent_C1	*E. faecium*	3.12	37.69	89	1.68	3472	3.9	96.1
ent_C11	*E. faecium*	2.74	37.94	5	2.28	2800	4.6	97.4
ent_C111	*E. italicus*	2.33	39.25	28	2.21	2332	3.5	97.5
ent_C116	*E. faecium*	2.79	37.87	9	2.26	2863	3.5	98.2
ent_C14	*E. faecium*	2.83	37.79	14	2.26	2874	3.2	98.7
ent_C143	*E. faecium*	2.67	37.89	70	2.11	2811	3.3	98.3
ent_C146	*E. faecium*	2.87	37.82	58	2.35	3010	3.4	95.0
ent_C151	*E. faecium*	2.78	37.91	36	2.41	2919	2.0	98.8
ent_C154	*E. faecium*	2.75	37.9	5	1.48	2793	3.9	96.7
ent_C155	*E. faecium*	2.85	37.86	34	0.88	2993	3.7	99.0
ent_C156	*E. faecium*	2.78	37.82	7	0.97	2840	3.0	96.1
ent_C157	*E. faecium*	2.79	37.87	8	2.59	2858	3.7	96.5
ent_C158	*E. faecalis*	2.90	37.45	13	2.85	2864	2.5	98.5
ent_C159	*E. faecium*	2.76	37.88	7	2.56	2812	2.1	98.4
ent_C179	*E. durans*	2.94	37.87	13	2.54	2933	4.9	96.9

Table 1. *Cont.*

Strain ID	Genus & Species	Genome Size (Mb)	GC Content (%)	No of Scaffolds	N50 (Mb)	No of CDSs	Contamination (%)	Completeness (%)
ent_C22	*E. faecium*	2.81	37.88	49	2.14	2951	3.6	97.2
ent_C25	*E. faecium*	2.79	37.83	17	1.47	2832	4.6	95.4
ent_C28	*E. faecalis*	2.95	37.42	18	2.86	2933	2.3	97.5
ent_C29	*E. faecium*	2.77	37.86	13	1.64	2803	4.5	95.7
ent_C3	*E. faecium*	2.76	37.88	8	2.57	2806	3.4	97.2
ent_C4	*E. faecium*	2.79	37.86	11	1.20	2867	3.1	98.7
ent_C5	*E. faecium*	2.75	37.9	6	1.48	2784	4.1	97.5
ent_C57	*E. faecium*	2.75	37.9	4	2.41	2793	2.9	98.7
ent_C6	*E. faecium*	2.79	37.87	7	2.59	2865	2.2	98.4
ent_C62	*E. faecium*	2.85	37.83	38	2.42	2897	2.2	95.8
ent_C7	*E. faecium*	2.79	37.87	12	1.63	2846	4.1	95.1
ent_C71	*E. faecium*	2.76	37.87	15	2.08	2795	3.7	98.4
ent_C74	*E. italicus*	2.39	39.36	29	1.43	2442	3.2	97.4
ent_C78	*E. faecalis*	2.96	37.46	18	2.43	2918	2.8	98.0
ent_C8	*E. faecium*	2.74	37.94	5	1.64	2798	3.2	98.9

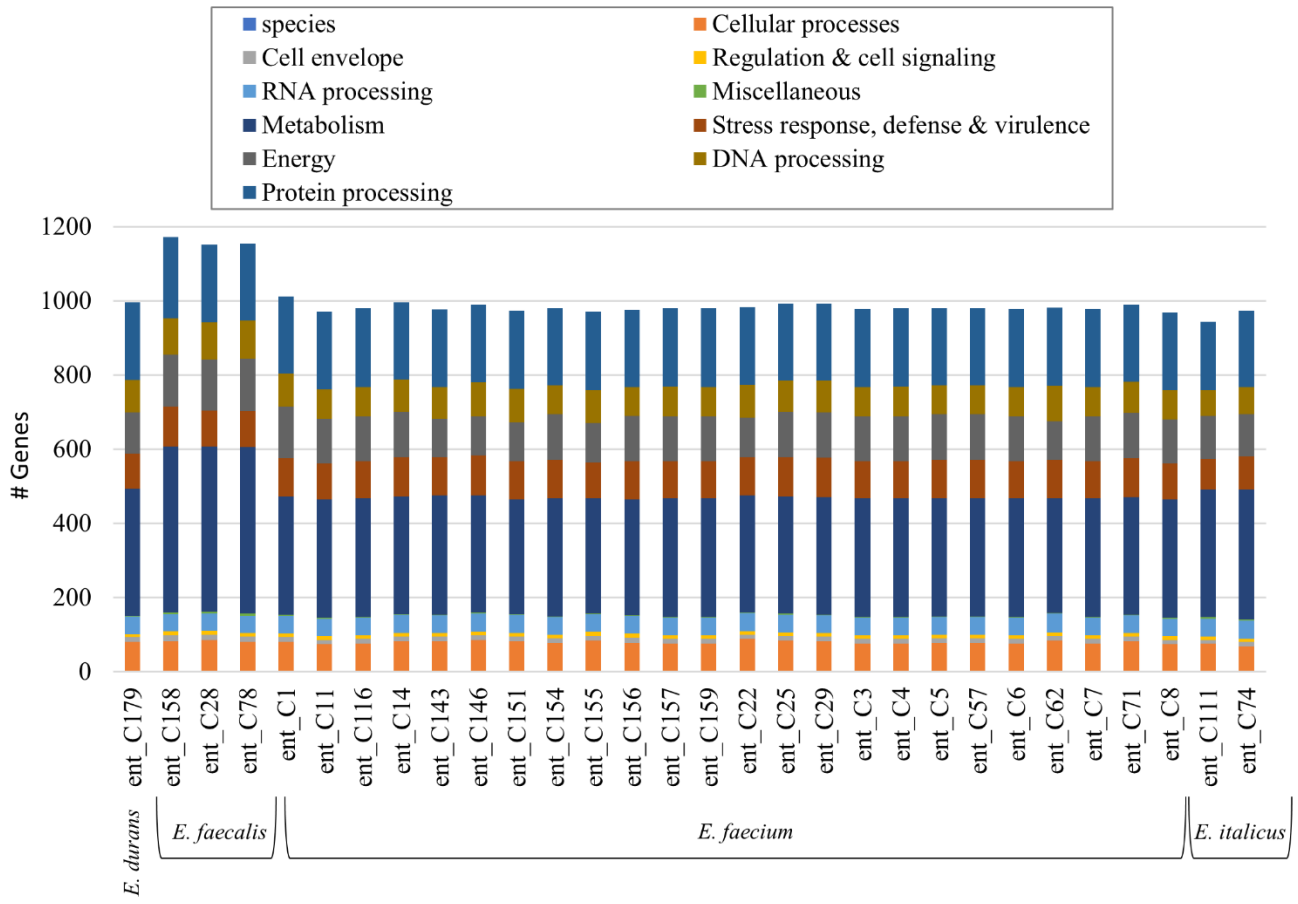


Figure 1. Overview of the subsystems in *Enterococcus* spp. genomes.

3.2. Presence of Resistance Genes

Analysis with Abricate and the Resfinder database revealed the presence of 124 RGs in the collection of *Enterococcus* spp., conferring resistance to aminoglycosides, macrolides, lincosamides and streptogramines (MLS), drug and biocide resistance (efflux pumps) and tetracyclines (Figure 2). Less frequent types of resistance were those to trimethoprim, nucleosides and of efflux pumps conferring multi-drug resistance (MDR). *Enterococcus* spp.

isolates carried four AMR genes, on average, while all (n = 30) isolates carried at least one AMR gene. Details on the RG-content per isolate are presented in Supplementary File S1. The most prevalent types of resistance were those to drug and biocide resistance and MLS, with 90.0% of isolates carrying at least one related gene, followed closely by aminoglycoside resistance (86.7%).

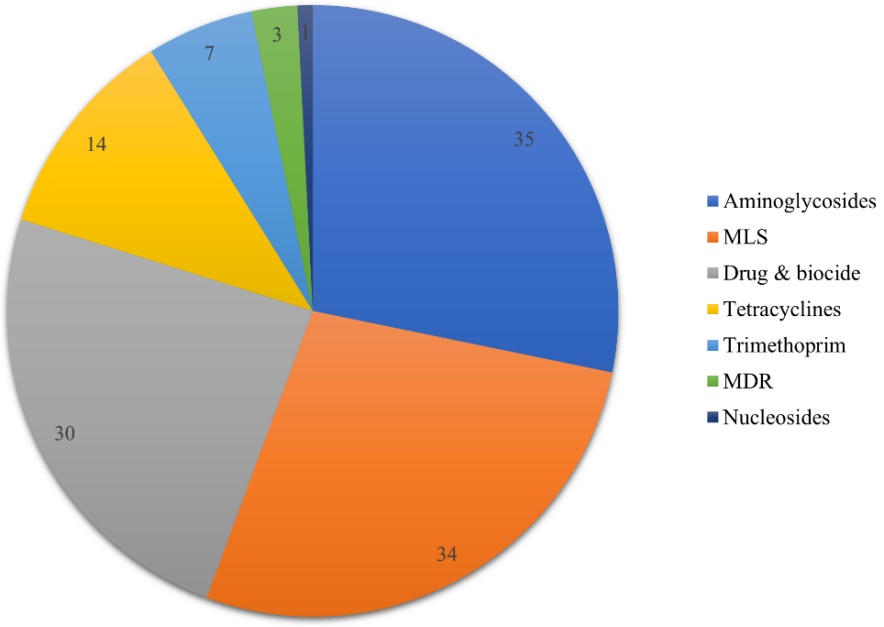


Figure 2. Distribution and number of AMR genes found in *Enterococcus* spp. per conferred antimicrobial class resistance.

Within the MLS category the *msrC* gene, encoding for an ABC efflux pump, was the most prevalent (71% of MLS genes), followed by *lsa* efflux pump (18%) and the *ermB* 23S methyltransferase (12%). The drug & biocide category was dominated by the *efmA* group efflux pumps. Acquired, high-level aminoglycoside resistance is particularly problematic in clinical Enterococci isolates [44] and within this category we found *aac(6′)-Im* in 83.3% of isolates. This acetyltransferase alters the 6-prime amino group of the aminoglycoside molecule by transferring an acetyl group from acetyl-CoA to the aminoglycoside drug and is the most common resistance type to aminoglycosides used in microbial chemotherapy [45]. The *ant(6)*, *ant(9)* nucleotidyltransferases and *aph(2′′)* phosphotransferase were also found but were less frequent. Tetracycline resistance genes were present in one-third of *Enterococcus* spp. More specifically, the *tetM* and *tetlL* genes, encoding tetracycline ribosomal protection proteins, were present in eight and six strains, respectively. Furthermore, trimethoprim resistance genes (*dfr*-like) were found in 23.3% of strains, whereas *lsaE* efflux pump gene, which can extrude a broad range of antimicrobials and confer an MDR phenotype in Enterococci that carry it [46], was found in three isolates. Lastly, the *sat* nucleoside was found in only one *E. faecalis* strain.

3.3. Presence of Virulence Genes

The virulence gene analysis revealed the presence of 94 VGs distributed in four major categories (Figure 3). All but one isolates (n = 29) carried at least one VG and three VGs on average. Details of VG-content per isolate are presented in Supplementary File S1. Adherence factors, mainly the *acm* and *scm* adhesins, were ubiquitous (100% of VG-carrying strains). The remaining VG categories had genes present only in three *E. faecalis* strains (ent_C28, ent_C78 and ent_C158). These stains carried the *csp* operon genes, encoding for the biosynthesis of capsular polysaccharides which contribute to host immune evasion, but only two out of three strains had the seven open reading frames in the *cps* operon that

are essential for capsule production (i.e., *cpsC*, *cpsD*, *cpsE*, *cpsG*, *cpsI*, *cpsJ*, and *cpsK*) [47,48]. Lastly, in the exoenzymes category, the *gelE* and *sprE* proteases, suggested to strongly contribute to pathogenesis in several infection models [49], were found in two out of the three *E. faecalis* strains. Moreover, none of the isolates in our collection contained the important exotoxins *cylR2*, *cylL-1*, *cylL-s*, or *cylM*.

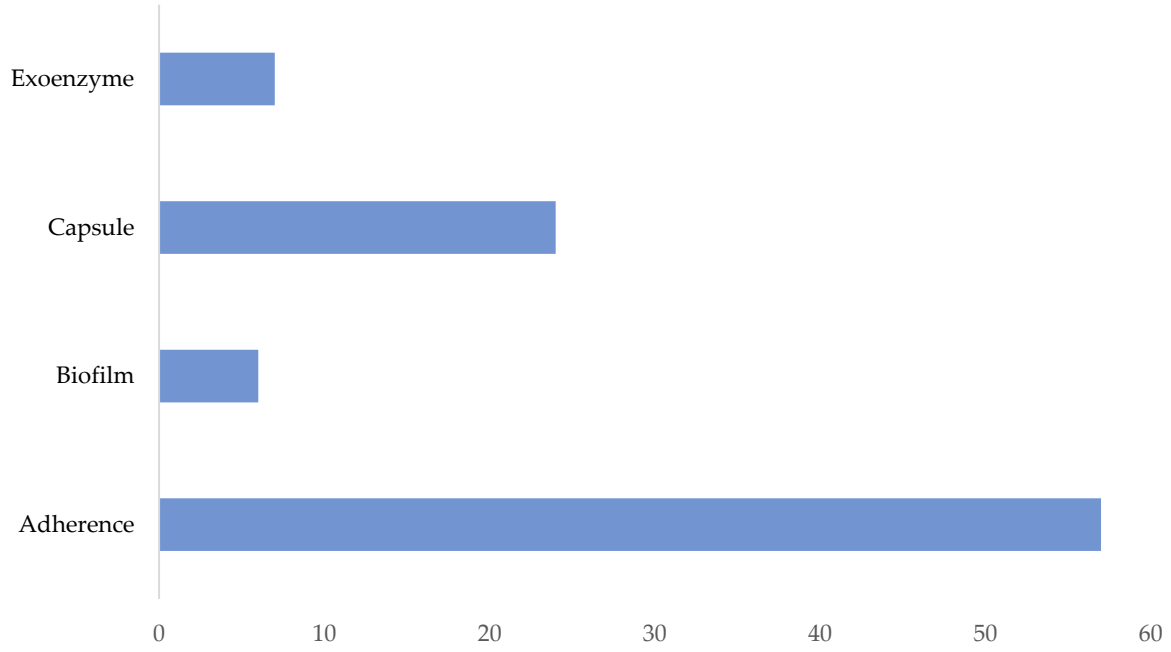


Figure 3. Distribution of virulence genes (VG) found in *Enterococcus* spp. per VG category function.

3.4. Other Genomic Features

3.4.1. Bacteriocins, Prophages and CRISPR-Cas

All isolates but two contained bacteriocins (n =28). The most prevalent bacteriocins were the enterocins (*ent*) A (n = 23), B (n = 20) and P (n = 24) as well as enterolysin A (n = 23). All other bacteriocins (e.g., Enterocin_SE-K4, UviB and Bac32) were each present in less than ten isolates. Moreover, 29 strains showed the presence of one or more complete prophage regions in their genomes with a maximum of three regions in four strains. One isolate of each of *E. faecalis*, *E. durans,* and *E. italicus* showed the presence CRISPR sequences (evidence level = 4) and *cas* genes in their genomes.

3.4.2. Plasmids and other MGEs

Detailed results of the 54 plasmids carried by *Enterococcus* spp. isolates are presented in Table S1. Enterococci harbored two plasmids, on average. The majority of plasmids (n = 16) were small, non-typable plasmids. The *Inc18* replicon type plasmids were present in 12 isolates (Table S1).

Ninety-seven MGEs were identified in the 30 Enterococci while isolates carried three MGEs on average. All the MGEs were insertion sequence (IS) elements with the most prevalent being IS*Enfa4* (18.6% of MGEs). IS*Efm2/ISLgar5* was the second most dominant IS (16.5%) followed by IS*Lpl1* (14.4%), an IS first described in *Lactobacillus plantarum*. All other IS elements made up less than 10% of the total MGEs.

Moreover, we used the PathogenFinder machine-learning algorithm to predict the pathogenicity of the isolates in our collection and thus classify them as human pathogens or commensals. Interestingly, all the isolates were predicted as potential human pathogens.

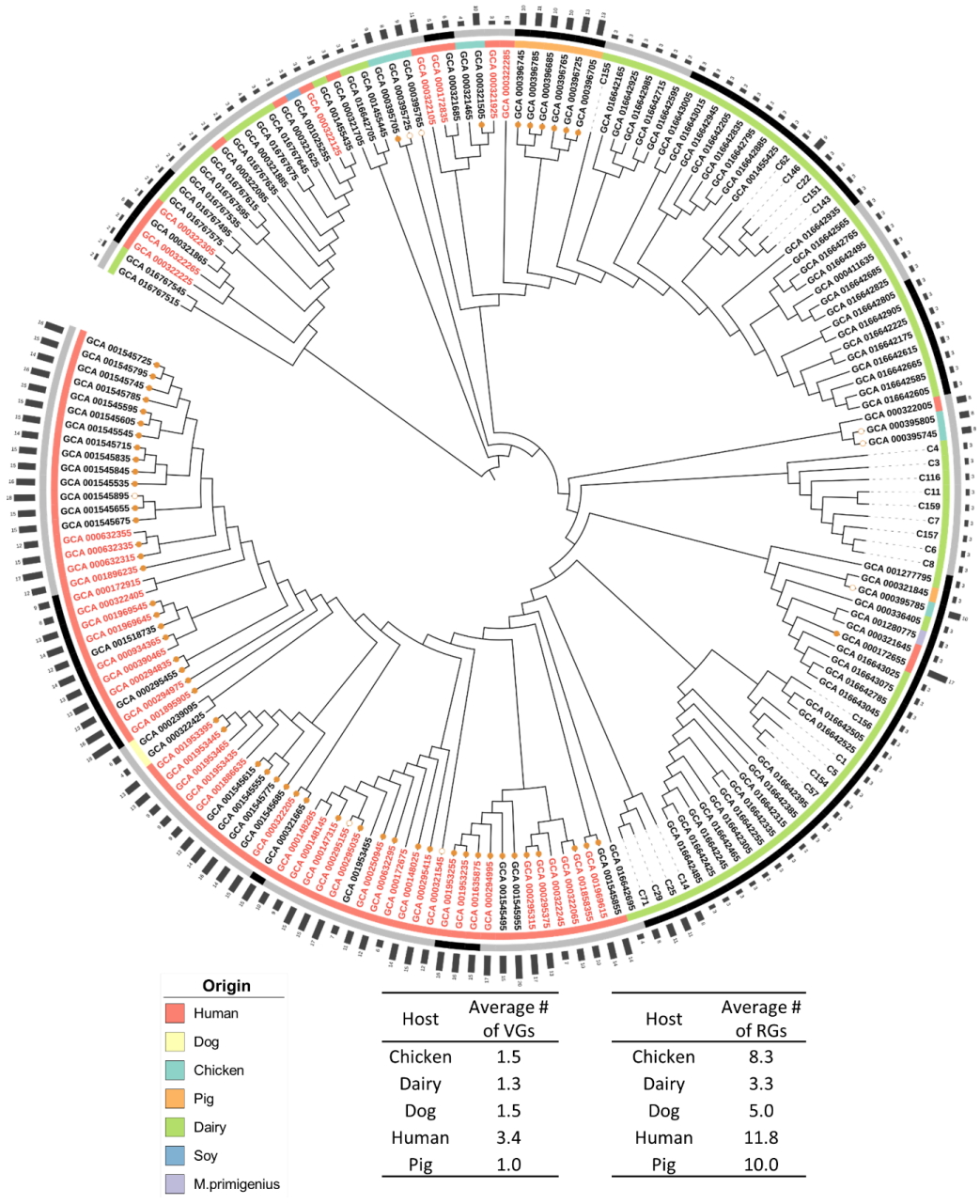
3.5. Phylogenetic Analysis and Comparative Genomics

In silico typing revealed the genetic background of the sequenced *E. faecium* (n = 24) and *E. faecalis* (n = 3) isolates by assignment to six and three unique sequence types (STs), respectively while for the other species (*E. durans* and *E. italicus*), no official MLST schemas exist (https://pubmlst.org/ accessed on 11 May 2022). For *E. faecium*, most of the isolates (nine in 24) belonged to ST2184, followed by ST545 (n = 5), ST1956 and ST113 (n = 4, each). The three *E. faecalis* isolates were assigned to ST19, ST205, and ST206. *E. faecium* ST113 and *E. faecalis* ST205 have been previously reported to be associated with infections [4,49].

Given that *E. faecium* was the dominant species isolated from the dairy products of this study, a comparative genomic analysis was conducted to gain deeper insights into the genetic relationships of the *E. faecium* isolated from various sources. More specifically, genomes of human (n = 81), dairy product (n = 61), chicken (n = 8), pig (n = 7), dog (n = 2), and soy (n = 1) origin as well as one isolate of *M. primigenius* were parsed from the NCBI database [50] (Figure 4), creating an extension of the dataset initially reported by Zhong et al. [51]. All animal isolates were retrieved from the gastrointestinal (GI) tract while those of human origin were isolated from bloodstream infections (n = 39/81), urinary tract infections (n = 9/81) and the GI tract (33 in 81) (Figure 4). The pangenome of the *E. faecium* isolates (i.e., the total number of unique gene families found) consisted of 15,951 gene families. The core-genome comprised 1327 genes, which is equivalent of 46.6% of the average number of CDSs in each genome, meaning that more than half of the protein-coding genes in each genome had an accessory role.

The phylogenetic tree and the hierarchical clusters (n = 16) calculated with hierBAPS revealed the genomic relatedness of the analyzed *E. faecium* isolates. Isolates clustered primarily according to their isolation origin (Figure 4). The bottom-left part of the tree predominantly comprised clusters that contained mainly to human-origin isolates, retrieved from the GI tract (human-GI) or from clinical cases (human-clinical). Conversely, at the right part of the tree, we found the dairy-origin isolates, including all isolates of this study, organized in six closely related clusters. This arrangement was disrupted by a few isolates from other sources, including three from human GI (Figure 4). Lastly, the top-left of the tree contained isolate clusters with a mixture of human (GI and clinical), dairy, chicken, and pig origin. Nevertheless, the dominant, origin-driven pattern of *E. faecium* clustering observed in our extensive analysis corroborates the findings from previous studies [8,51,52], which also showed that human pathogenic isolates are generally phylogenetically distinct from environmental and food-related isolates.

One additional finding of our analysis, which was visually observable in the annotated tree, was the difference in terms of total RG-content between different origins, with human and pig isolates showing higher numbers of resistance determinants and strong association with vancomycin-resistance genes (Figure 4). We confirmed this pattern statistically with Scoary and found significant differences in the total number of RGs and VGs between source of isolation. More specifically, dairy isolates carried the lowest number of RGs (Figure 4) and their RG content was significantly different ($p < 0.05$) from all other sources except for dog isolates. Moreover, the RG analysis showed that the *van*-like resistance and regulator genes, which respectively encode and regulate resistance to vancomycin and other glycopeptides, were predominant in human *E. faecium* isolates and even more abundant in isolates from human bloodstream infections and UTIs (Figure 4). Regarding the VGs, significant differences ($p < 0.05$) were found only for the human VG-content versus all other origins (Figure 4).



Host	Average # of VGs
Chicken	1.5
Dairy	1.3
Dog	1.5
Human	3.4
Pig	1.0

Host	Average # of RGs
Chicken	8.3
Dairy	3.3
Dog	5.0
Human	11.8
Pig	10.0

Figure 4. Phylogenetic tree including the 24 *E. faecium* dairy isolates of this study and 161 isolates from various origins, parsed from NCBI. The outer bar graph denotes the abundance of RGs found in each strain. Alternating dark and light grey segments in the outer ring mark different hierBAPS clusters. The colored inner ring indicates the isolates' origin according to the legend. Isolate labels with red font signify human clinical isolates form cases of bloodstream and urinary tract infections. The outlined, filled or absent orange circle at the tree's leaves, shows whether an isolate carried vancomycin-resistance regulator gene(s) (*vanR*-like) only, regulator and vancomycin-resistance gene(s) (*van*-like), or no vancomycin-related genes, respectively. The isolate labels of this study and of those retrieved from NCBI have a "C" and "GCA_" numerical prefix, respectively. #, number of the respective genes.

In the next analysis, we aimed to juxtapose isolates of human and dairy origin with respect to their functional (CAZymes content) and safety (RGs-VGs) profile and elucidate whether the presence-absence patterns of these genes can distinguish isolates of different sources. The heatmap and hierarchical clustering showed an arrangement similar to that of

the phylogenetic tree. Three major clusters were formed, which predominantly contained dairy and human isolates (clinical and GI) and a mixture of isolates, respectively (Figure 5). Statistical analysis for gene enrichment with Scoary showed no significant associations of dairy isolates with specific RGs or VGs, on the contrary, several genes, such as the *van*-like RGs and the *ecbA* VG, showed a strong negative association (odds ratio (OR) < 1, *p*-value < 0.05). Conversely, several RGs (n = 19) and VGs (n = 4) were significantly enriched in the human-clinical (n = 13) and human-GI (n = 16) sources, while six genes were associated with both sources (OR > 1, *p*-value < 0.05). Important RGs enriched in the human sources were the vancomycin RGs (*van*-like) and regulator genes (*vanR*-like), the tetracycline RGs (*tet*-like), and the *lsaE* MDR efflux pump. Noteworthy VGs were the *sgrA* and *ecbA* adhesins as well as the *scm* adhesin. Several RGs and VGs, such as the *efmA* efflux pump and *acm* adhesin, were ubiquitous across all isolate origins (Figure 4).

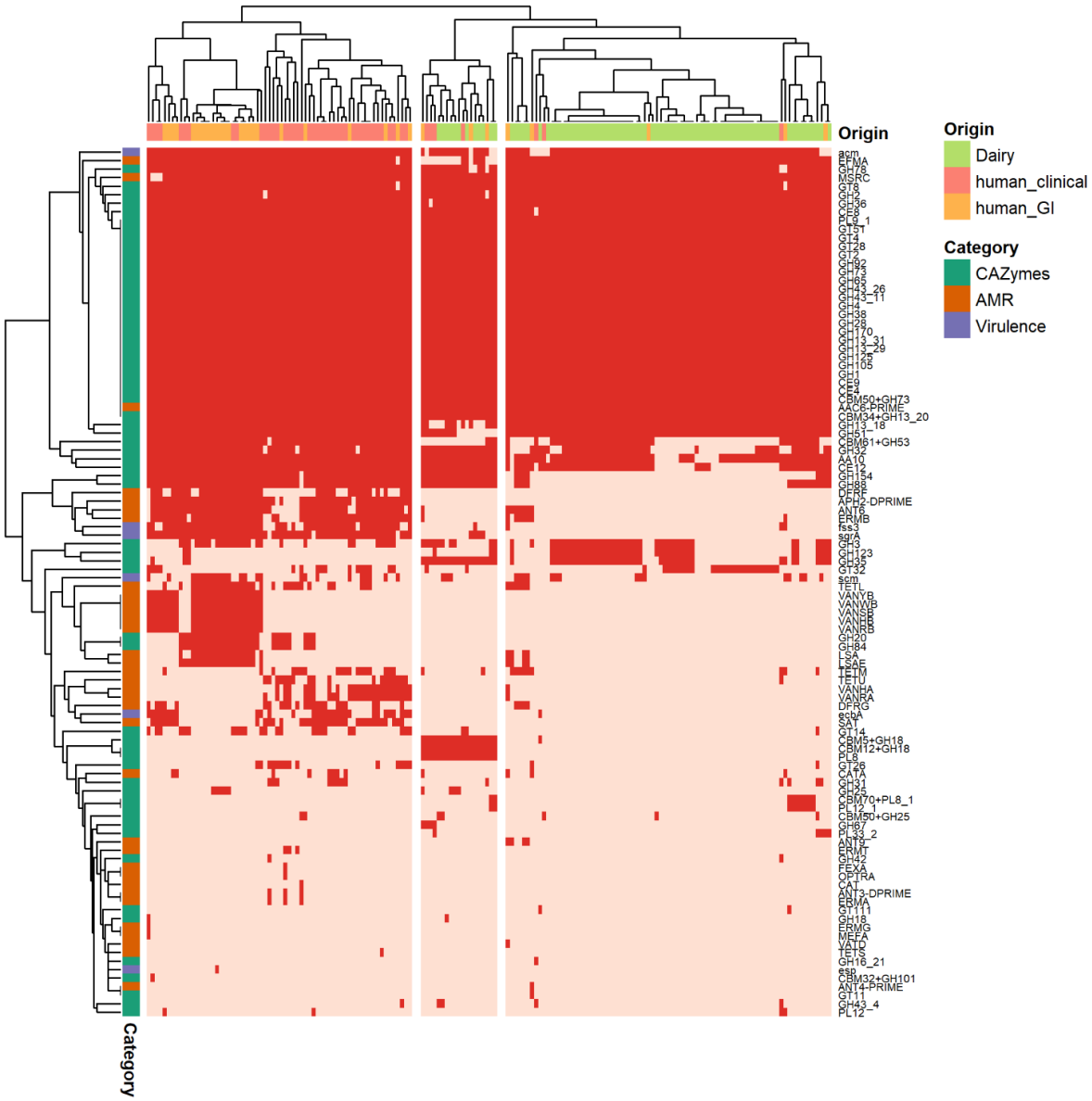


Figure 5. Cluster heatmap generated using a resistance, virulence and CAZyme gene presence-absence data matrix of *E. faecium* isolated from dairy, human-clinical and human-gastrointestinal origin.

The majority of the CAZymes identified in the analyzed *E. faecium* isolates belonged to the glycoside hydrolase (GH) families with the GH43 family being predominant, while

62 different families were identified. The GH43 family as well as others from different groups, e.g., Glycosyltransferase (GT) 2, Carbohydrate Esterase (CE) 4 and Polysaccharide Lyase (PL) 9, were ubiquitous among isolates. Nevertheless, our comparative analysis showed that several CAZyme families were associated with either the human or dairy sources. The GH35, GH123, and the Carbohydrate-binding module (CBM) 70 families were significantly enriched in dairy-related isolates (OR > 1, p-value < 0.05). Within the human sources (clinical & GI), the CBM61, CE12, and GH154 CAZymes were significantly enriched and they showed strong negative association in the dairy-origin isolates. The AA10 CAZyme was strongly associated only with the human–clinical source.

Lastly, analysis with Traitar for predicted phenotypic characteristics showed that, irrespective of their origin, all isolates can utilize sugars, such as lactose, galactose, maltose, sucrose, and D-mannitol (Figure 6). The clustered heatmap of predicted traits indicated an overlap of dairy and human origin isolates as well as a separate human–clinical cluster. Statistical analysis for phenotype enrichment provided more insights for this arrangement. Dairy-origin isolates had a strong association (OR > 1, p-value < 0.05) with the utilization of starch and malonate, whereas human origin isolates were found to be significantly related with the catabolism of L-rhamnose, D-sorbitol, and raffinose.

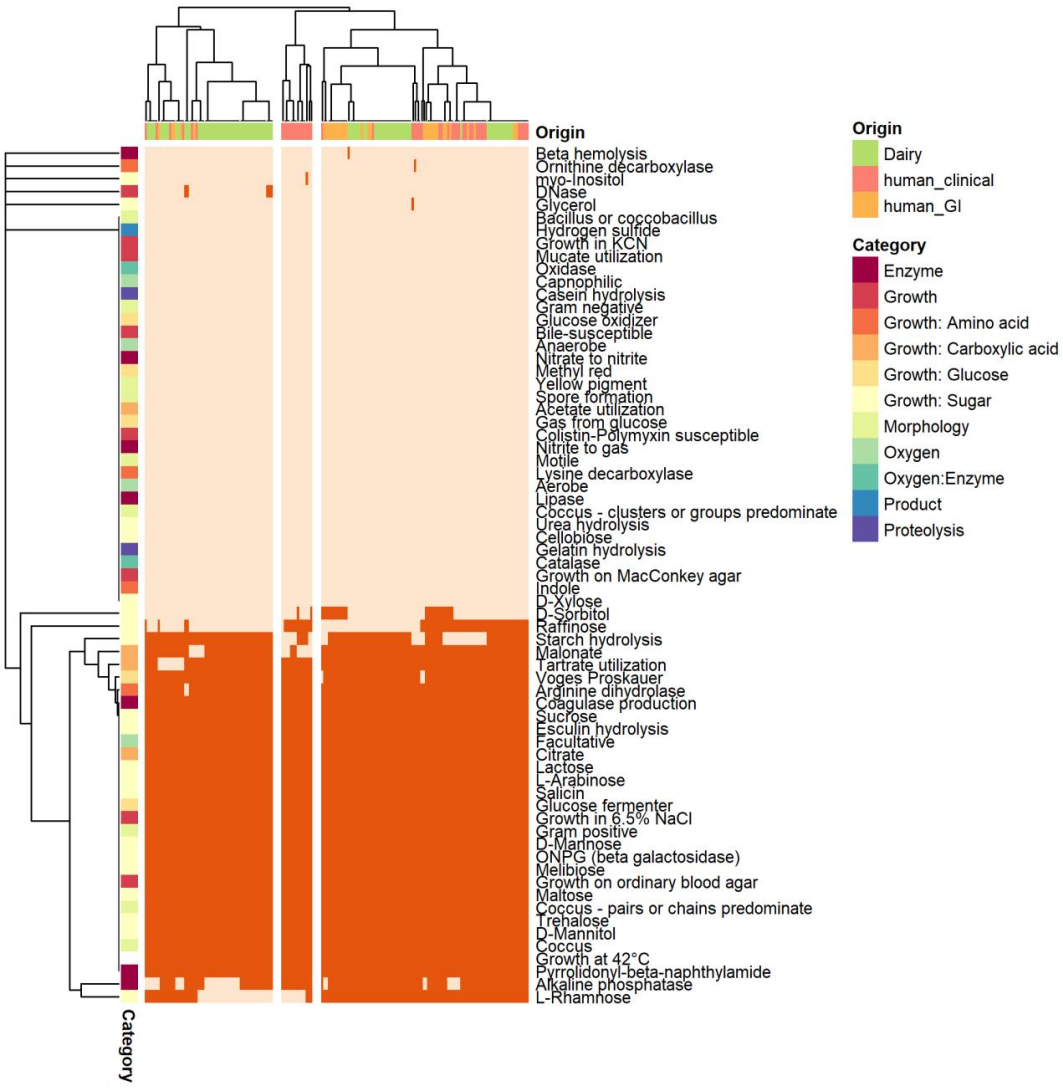


Figure 6. Predicted phenotypic characteristics of *E. faecium* isolates with Traitar.

4. Discussion

Enterococci are largely commensal organisms and several strains have been proven to have probiotic effects for their hosts. Furthermore, they can be used as adjunct cultures in fermenting dairy products, where they modulate the metabolic network created during fermentation and thus affect the organoleptic properties of the end products [8,9]. The dark side of Enterococci is manifested in several strains, mainly of *E. faecium* and *E. faecalis*, which are causative agents of persistent nosocomial infections. For this reason, *Enterococcus* species have not been granted the QPS status [9] and thus cannot be freely used as starter, adjunct cultures or generally as food-additives. In this context, we aimed to provide a high-resolution genomic characterization of an Enterococci collection retrieved from dairy products, such as raw sheep milk, artisanal Feta cheese, and artisanal Kefalograviera cheese, with *E. faecium* being the dominant isolated species and *E. faecalis*, *E. italicus*, and *E. durans* appearing less frequently.

A subsystem is a set of CDSs that together implement a specific biological process or structural complex [53]. Apart from an enriched metabolism repertoire in *E. faecalis*, the subsystem analysis did not reveal major differences in the main cellular processes of this Enterococci collection. Bacteriocins are peptides synthesized by bacteria that kill other bacteria by creating pores in target membranes. They have been suggested to play a promising role as food preservatives by modulating the food microbiome and suppressing pathogens [54]. Prophages that integrate in bacterial genomes often harbor resistance or virulence determinants [55]. Thus, the existence of CRISPR/Cas systems can help to protect bacterial genomes from prophage integration [8]. The abundance of bacteriocins and CRISPR/Cas systems suggested the ability of the isolates to suppress pathogens and evade bacteriophages, respectively. Furthermore, analysis for resistance and virulence determinants showed a diverse collection RGs and VGs present in Enterococci, though known hallmark pathogenicity factors were either absent, e.g., vancomycin RGs and the CylR2, CylL-1, CylL-s, and CylM exotoxins, or occurred rarely, e.g., the *gelE* VG found in two out of 30 isolates. It is important to note, however, that the absence of hallmark pathogenicity factors cannot guarantee the safety of isolates. With regard to the plasmids and other MGEs, we found the Inc18 replicon type plasmids to be prevalent. These plasmids have been found to carry genes that encode resistance to vancomycin, chloramphenicol, as well as the MLS group of antibiotics. They are also widespread in nosocomial settings, but they have been isolated from food-producing animals, the environment, and wastewater facilities as well [56]. The most prevalent IS element was IS*Enfa4*. Although this IS has been associated with the MDR gene *cfr* [57], this gene was not identified in our collection. IS*Efm2/ISLgar5*, the second most dominant IS, has been linked to the *Tn1546*-like transposon that often harbors the *vanA* vancomycin RG [57]. However, none of the two genetic elements were detected in our collection. Furthermore, IS*Lpl1* (14.4%), an IS first described in *Lactobacillus plantarum*, is considered to be linked to bacteriocin production and sugar utilization [58,59].

Interestingly, all the isolates were predicted as potential human pathogens by the PathogenFinder machine learning algorithm. However, in their publication, Cosentino et al. [34] report that the performance of their classifier was suboptimal for Enterococci, explained by the few *Enterococcus* spp. isolates included in their training dataset. We thus focused on the comparative genomic and phylogenetic analysis to juxtapose de facto pathogenic isolates from human infections and isolates from dairy and other sources.

The comparative genomic analysis with a large collection of *E. faecium* isolates from various origins provided deeper insights. Although phylogenetically diverse, clusters of closely related isolates were formed according to their origin, with three large dairy, human, and mixed cluster arrangements appearing in the phylogenetic tree (Figure 4). A similar structure was observed in the hierarchical cluster-heatmap created with the resistome, virulome, and CAZyme content of these isolates (Figure 5). Statistical analysis using the presence–absence of related gene suggested that *E. faecium* of dairy origin are not associated with particular factors, while those of human origin (both clinical and intestinal)

are strongly linked with notorious resistance and virulence determinants. Important RGs enriched in the human sources were the vancomycin RGs (*van*-like) and regulator genes (*vanR*-like), the tetracycline RGs (*tet*-like), and the *lsaE* MDR efflux pump. These RGs, especially the vancomycin resistance determinants, are strongly associated with persistent nosocomial infections, which alarmingly show increasing rates and are becoming more difficult to manage [5,6,52]. Noteworthy VGs were the *sgrA* and *ecbA* adhesins, which were shown to play a role in adhesion to medical-device-related infections by forming biofilms [60], as well as the *scm* adhesin. Several RGs and VGs, such as the *efmA* efflux pump and *acm* adhesin, were ubiquitous across all isolate origins (Figure 4). This finding suggests that certain genetic elements might not be directly linked to pathogenicity and, in the absence of other hallmark pathogenic factors, they might contribute to the attachment and survival of Enterococci in the GI tract where the bacteria can exert their beneficial probiotic properties [7,8,61].

Enzymes responsible for the synthesis and metabolism of carbohydrates are known as carbohydrate-active enzymes (CAZymes). Apart from being interesting in biotechnological applications, the biotransformation of food carbohydrates by bacteria can produce valuable metabolites [62]. Additionally, the combination of pre- and probiotics can lead to significant beneficial effects such as the inhibition of inflammatory processes and the reduction of cholesterol levels [63]. CBMs were significantly enriched in dairy-related isolates. They act as catalytic modules of long CAZymes, such as glycoside hydrolases, with the latter being essential in the degradation of complex carbohydrates, such as lactose and starch [64]. Within the human sources, the CBM61, CE12, and GH154 CAZymes were significantly enriched, which also showed a strong negative association in the dairy origin isolates. Interestingly, the AA10 CAZyme, which was strongly associated only with the human–clinical source, was found to play a key role in the binding of (opportunistic) pathogens [65,66].

Lastly, analysis with Traitar highlighted associations between the isolate source and specific phenotypic traits. Remarkably, dairy origin isolates presented a strong association with the utilization of starch and malonate, suggesting that these strains could be used in both dairy and vegetable fermentation [67]. On the contrary, human origin isolates were found to be significantly related with the catabolism of L-rhamnose, D-sorbitol, and raffinose. The fermentation of the latter two sugars has indeed been found to be associated with nosocomial *E. faecium* pathogens [68,69].

Even though our comparative analysis provided some notable insights regarding the genomic composition of Enterococci in the context of their origin, our approach cannot provide a definitive answer as to which genomic features can distinguish pathogenic from commensal isolates. Although beyond the scope of this study, this goal can be achieved through a systematic genome-wide association study (GWAS) coupled with machine learning methods and involving large numbers (1000–2000) of properly labeled isolates (e.g., pathogenic vs. commensal) to establish sufficient statistical power [70]. Such an effort should involve modelling based on all core genomic features (e.g., core-genes and their SNPs), given that the pathogenic lifestyle of Enterococci is likely explained by the interplay of multiple genetic factors. To distinguish friend from foe in Enterococci and exploit the beneficial properties of harmless strains, future projects should aim to accomplish this difficult objective.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/app12199620/s1, Table S1: Plasmids present in *Enterococcus* spp. isolates and their characteristics, File S1: Details on the RG-content per isolate.

Author Contributions: Conceptualization, M.M.; methodology, I.A. and M.M.; software, I.A.; formal analysis, I.A. and M.M.; investigation, I.A. and M.T.; resources, M.M.; data curation, I.A. and M.M.; writing—original draft preparation, I.A.; writing—review and editing, I.A., M.T., S.P. and M.M.; supervision, S.P. and M.M.; project administration, M.M.; funding acquisition, M.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship, and Innovation, under the call RESEARCH—CREATE—INNOVATE (T1EDK-02087).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: This Whole Genome Shotgun project (BioProject number PRJNA847029) has been deposited at DDBJ/ENA/GenBank under the accession numbers JAMRUQ000000000 to JAMRUZ000000000 and JAMRVA000000000 to JAMRVT000000000. The version described in this paper is version JAMRUQ010000000 to JAMRUZ010000000 and JAMRVA010000000 to JAMRVT010000000.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Różańska, H.; Lewtak-Piłat, A.; Kubajka, M.; Weiner, M. Occurrence of Enterococci in Mastitic Cow's Milk and Their Antimicrobial Resistance. *J. Vet. Res.* **2019**, *63*, 93–97. [CrossRef] [PubMed]
2. Jan, O.; Jaśkowski, J.M. Mastitis in Small Ruminants. *Med. Weter.* **2014**, *70*, 67–72. [CrossRef]
3. Bergonier, D.; de Crémoux, R.; Rupp, R.; Lagriffoul, G.; Berthelot, X. Mastitis of Dairy Small Ruminants. *Vet. Res.* **2003**, *34*, 689–716. [CrossRef] [PubMed]
4. Hammerum, A.M. Enterococci of Animal Origin and Their Significance for Public Health. *Clin. Microbiol. Infect.* **2012**, *18*, 619–625. [CrossRef]
5. Markwart, R.; Willrich, N.; Haller, S.; Noll, I.; Koppe, U.; Werner, G.; Eckmanns, T.; Reuss, A. The Rise in Vancomy-Cin-Resistant *Enterococcus faecium* in Germany: Data from the German Antimicrobial Resistance Surveillance (ARS). *Antimicrob. Resist. Infect. Control.* **2019**, *8*, 1–11. [CrossRef]
6. Hammerum, A.M.; Lester, C.H.; Heuer, O.E. Antimicrobial-Resistant Enterococci in Animals and Meat: A Human Health Hazard? *Foodborne Pathog. Dis.* **2010**, *7*, 1137–1146. [CrossRef]
7. Dapkevicius, M.; Sgardioli, B.; Câmara, S.; Poeta, P.; Malcata, F. Current Trends of Enterococci in Dairy Products: A Comprehensive Review of Their Multiple Roles. *Foods* **2021**, *10*, 821. [CrossRef]
8. Ghattargi, V.C.; Gaikwad, M.A.; Meti, B.S.; Nimonkar, Y.S.; Dixit, K.; Prakash, O.; Shouche, Y.S.; Pawar, S.P.; Dhotre, D.P. Comparative Genome Analysis Reveals Key Genetic Factors Associated with Probiotic Property in *Enterococcus faecium* Strains. *BMC Genom.* **2018**, *19*, 652. [CrossRef]
9. Graham, K.; Stack, H.; Rea, R. Safety, Beneficial and Technological Properties of Enterococci for Use in Functional Food Applications—A Review. *Crit. Rev. Food Sci. Nutr.* **2020**, *60*, 3836–3861. [CrossRef]
10. Tsigkrimani, M.; Bakogianni, M.; Paramithiotis, S.; Bosnea, L.; Pappa, E.; Drosinos, E.H.; Skandamis, P.N.; Mataragas, M. Microbial Ecology of Artisanal Feta and Kefalograviera Cheeses, Part I: Bacterial Community and Its Functional Characteristics with Focus on Lactic Acid Bacteria as Determined by Culture-Dependent Methods and Phenotype Microarrays. *Microorganisms* **2022**, *10*, 161. [CrossRef]
11. Tsigkrimani, M.; Panagiotarea, K.; Paramithiotis, S.; Bosnea, L.; Pappa, E.; Drosinos, E.H.; Skandamis, P.N.; Mataragas, M. Microbial Ecology of Sheep Milk, Artisanal Feta, and Kefalograviera Cheeses. Part II: Technological, Safety, and Probiotic Attributes of Lactic Acid Bacteria Isolates. *Foods* **2022**, *11*, 459. [CrossRef] [PubMed]
12. Syrokou, M.K.; Themeli, C.; Paramithiotis, S.; Mataragas, M.; Bosnea, L.; Argyri, A.A.; Chorianopoulos, N.G.; Skandamis, P.N.; Drosinos, E.H. Microbial Ecology of Greek Wheat Sourdoughs, Identified by a Culture-Dependent and a Culture-Independent Approach. *Foods* **2020**, *9*, 1603. [CrossRef] [PubMed]
13. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2019. Available online: http://www.bioinformatics.babraham.ac.uk/projects/fastqc (accessed on 11 May 2022).
14. Arkin, A.P.; Cottingham, R.W.; Henry, C.S.; Harris, N.L.; Stevens, R.L.; Maslov, S.; Dehal, P.; Ware, D.; Perez, F.; Canon, S.; et al. KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nat. Biotechnol.* **2018**, *36*, 566–569. [CrossRef]
15. Wick, R.R.; Judd, L.M.; Gorrie, C.L.; Holt, K.E. Unicycler: Resolving Bacterial Genome Assemblies from Short and Long Sequencing Reads. *PLoS Comput. Biol.* **2017**, *13*, e1005595. [CrossRef] [PubMed]
16. Davis, J.J.; Wattam, A.R.; Aziz, R.K.; Brettin, T.; Butler, R.; Butler, R.M.; Chlenski, P.; Conrad, N.; Dickerman, A.; Dietrich, E.M.; et al. The PATRIC Bioinformatics Resource Center: Expanding Data and Analysis Capabilities. *Nucleic Acids Res.* **2020**, *48*, D606–D612. [CrossRef]
17. Walker, B.J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.; Wortman, J.; Young, S.K.; et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS ONE* **2014**, *9*, e112963. [CrossRef] [PubMed]
18. Bosi, E.; Donati, B.; Galardini, M.; Brunetti, S.; Sagot, M.-F.; Lio, P.; Crescenzi, P.; Fani, R.; Fondi, M. MeDuSa: A Multi-Draft Based Scaffolder. *Bioinformatics* **2015**, *31*, 2443–2451. [CrossRef] [PubMed]

19. Parks, D.H.; Imelfort, M.; Skennerton, C.T.; Hugenholtz, P.; Tyson, G.W. CheckM: Assessing the Quality of Microbial Genomes Recovered from Isolates, Single Cells, and Metagenomes. *Genome Res.* **2015**, *25*, 1043–1055. [CrossRef]
20. Lu, J.; Salzberg, S.L. SkewIT: The Skew Index Test for Large-Scale GC Skew Analysis of Bacterial Genomes. *PLoS Comput. Biol.* **2020**, *16*, e1008439. [CrossRef]
21. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality Assessment Tool for Genome Assemblies. *Bioinformatics* **2013**, *29*, 1072–1075. [CrossRef]
22. Wood, D.E.; Lu, J.; Langmead, B. Improved Metagenomic Analysis with Kraken 2. *Genome Biol.* **2019**, *20*, 257. [CrossRef] [PubMed]
23. Meier-Kolthoff, J.P.; Göker, M. TYGS is an Automated High-Throughput Platform for State-of-the-Art Genome-Based Taxonomy. *Nat. Commun.* **2019**, *10*, 1–10. [CrossRef]
24. Seemann, T. Prokka: Rapid Prokaryotic Genome Annotation. *Bioinformatics* **2014**, *30*, 2068–2069. [CrossRef]
25. Tatusov, R.L.; Fedorova, N.D.; Jackson, J.D.; Jacobs, A.R.; Kiryutin, B.; Koonin, E.V.; Krylov, D.M.; Mazumder, R.; Smirnov, S.; Nikolskaya, A.N.; et al. The COG Database: An Updated Vesion Includes Eukaryotes. *BMC Bioinform.* **2003**, *4*, 1–14. [CrossRef]
26. Couvin, D.; Bernheim, A.; Toffano-Nioche, C.; Touchon, M.; Michalik, J.; Néron, B.; Rocha, E.P.C.; Vergnaud, G.; Gautheret, D.; Pourcel, C. CRISPRCasFinder, An Update of CRISRFinder, Includes A Portable Version, Enhanced Performance and Integrates Search for Cas Proteins. *Nucleic Acids Res.* **2018**, *46*, W246–W251. [CrossRef] [PubMed]
27. Arndt, D.; Marcu, A.; Liang, Y.; Wishart, D.S. PHAST, PHASTER and PHASTEST: Tools for Finding Prophage in Bacterial Genomes. *Briefings Bioinform.* **2017**, *20*, 1560–1567. [CrossRef] [PubMed]
28. Seemann, T. *Abricate, Github*; Github. 2020. Available online: https://github.com/tseemann/abricate (accessed on 11 May 2022).
29. Zankari, E.; Hasman, H.; Cosentino, S.; Vestergaard, M.; Rasmussen, S.; Lund, O.; Aarestrup, F.M.; Larsen, M.V. Identification of Acquired Antimicrobial Resistance Genes. *J. Antimicrob. Chemother.* **2012**, *67*, 2640–2644. [CrossRef]
30. Chen, L.; Zheng, D.; Liu, B.; Yang, J.; Jin, Q. VFDB 2016: Hierarchical and Refined Dataset for Big Data Analysis—10 Years on. *Nucleic Acids Res.* **2015**, *44*, D694–D697. [CrossRef]
31. Johansson, M.H.K.; Bortolaia, V.; Tansirichaiya, S.; Aarestrup, F.M.; Roberts, A.P.; Petersen, T.N. Detection of Mobile Genetic Elements Associated with Antibiotic Resistance in *Salmonella enterica* Using a Newly Developed Web Tool: MobileElementFinder. *J. Antimicrob. Chemother.* **2020**, *76*, 101–109. [CrossRef]
32. Van Heel, A.J.; De Jong, A.; Song, C.; Viel, J.H.; Kok, J.; Kuipers, O.P. BAGEL4: A User-Friendly Web Server to Thoroughly Mine Ripps and Bacteriocins. *Nucleic Acids Res.* **2018**, *46*, W278–W281. [CrossRef]
33. Robertson, J.; Nash, J.H.E. MOB-Suite: Software Tools for Clustering, Reconstruction and Typing of Plasmids from Draft Assemblies. *Microb. Genom.* **2018**, *4*, e000206. [CrossRef] [PubMed]
34. Cosentino, S.; Voldby Larsen, M.; Møller Aarestrup, F.; Lund, O. PathogenFinder—Distinguishing Friend from Foe Using Bacterial Whole Genome Sequence Data. *PLoS ONE* **2013**, *8*, e77302. [CrossRef]
35. 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*, 3691–3693. [CrossRef]
36. Croucher, N.J.; Page, A.J.; Connor, T.R.; Delaney, A.J.; Keane, J.A.; Bentley, S.D.; Parkhill, J.; Harris, S.R. Rapid Phylogenetic Analysis of Large Samples of Recombinant Bacterial Whole Genome Sequences Using Gubbins. *Nucleic Acids Res.* **2015**, *43*, e15. [CrossRef]
37. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. *Mol. Biol. Evol.* **2009**, *26*, 1641–1650. [CrossRef] [PubMed]
38. Cheng, L.; Connor, T.; Sirén, J.; Aanensen, D.M.; Corander, J. Hierarchical and Spatially Explicit Clustering of DNA Sequences with BAPS Software. *Mol. Biol. Evol.* **2013**, *30*, 1224–1228. [CrossRef]
39. Letunic, I.; Bork, P. Interactive Tree of Life (iTOL) v5: An Online Tool for Phylogenetic Tree Display and Annotation. *Nucleic Acids Res.* **2021**, *49*, W293–W296. [CrossRef]
40. Zhang, H.; Yohe, T.; Huang, L.; Entwistle, S.; Wu, P.; Yang, Z.; Busk, P.K.; Xu, Y.; Yin, Y. dbCAN2: A Meta Server for Automated Carbohydrate-Active Enzyme Annotation. *Nucleic Acids Res.* **2018**, *46*, W95–W101. [CrossRef]
41. Weimann, A.; Mooren, K.; Frank, J.; Pope, P.B.; Bremges, A.; McHardy, A.C. From Genomes to Phenotypes: Traitar, the Microbial Trait Analyzer. *mSystems* **2016**, *1*, e00101-16. [CrossRef]
42. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020; Available online: https://www.R-project.org/ (accessed on 11 May 2022).
43. Brynildsrud, O.; Bohlin, J.; Scheffer, L.; Eldholm, V. Rapid Scoring of Genes in Microbial Pan-Genome-Wide Association Studies with Scoary. *Genome Biol.* **2016**, *17*, 238. [CrossRef]
44. Chow, J.W. Aminoglycoside Resistance in Enterococci. *Clin. Infect. Dis.* **2000**, *31*, 586–589. [CrossRef] [PubMed]
45. Vetting, M.W.; Park, C.H.; Hegde, S.S.; Jacoby, G.A.; Hooper, D.C.; Blanchard, J.S. Mechanistic and Structural Analysis of Aminoglycoside *N*-Acetyltransferase AAC(6′)-Ib and Its Bifunctional, Fluoroquinolone-Active AAC(6′)-Ib-cr Variant. *Biochemistry* **2008**, *47*, 9825–9835. [CrossRef] [PubMed]
46. Salah, A.N.; Elleboudy, N.S.; El-Housseiny, G.S.; Yassien, M.A. Cloning and Sequencing of Lsae Efflux Pump Gene from MDR Enterococci and Its Role in Erythromycin Resistance. *Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis.* **2021**, *94*, 105010. [CrossRef] [PubMed]
47. Hancock, L.E.; Gilmore, M.S. The Capsular Polysaccharide of *Enterococcus faecalis* and Its Relationship to Other Polysaccharides in the Cell Wall. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 1574–1579. [CrossRef]

48. Thurlow, L.R.; Thomas, V.C.; Fleming, S.D.; Hancock, L.E. *Enterococcus faecalis* Capsular Polysaccharide Serotypes C and D and Their Contributions to Host Innate Immune Evasion. *Infect. Immun.* **2009**, *77*, 5551–5557. [CrossRef]
49. Engelbert, M.; Mylonakis, E.; Ausubel, F.M.; Calderwood, S.B.; Gilmore, M.S. Contribution of Gelatinase, Serine Protease, and *fsr* to the Pathogenesis of *Enterococcus faecalis* Endophthalmitis. *Infect. Immun.* **2004**, *72*, 3628–3633. [CrossRef]
50. O'Leary, N.A.; Wright, M.W.; Brister, J.R.; Ciufo, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.; et al. Reference Sequence (RefSeq) Database at NCBI: Current Status, Taxonomic Expansion, and Functional Annotation. *Nucleic Acids Res.* **2016**, *44*, D733–D745. [CrossRef]
51. Zhong, Z.; Kwok, L.-Y.; Hou, Q.; Sun, Y.; Li, W.; Zhang, H.; Sun, Z. Comparative Genomic Analysis Revealed Great Plasticity and Environmental Adaptation of the Genomes of *Enterococcus faecium*. *BMC Genom.* **2019**, *20*, 1–13. [CrossRef]
52. 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*, e00534-13. [CrossRef]
53. Overbeek, R.; Begley, T.; Butler, R.M.; Choudhuri, J.V.; Chuang, H.Y.; Cohoon, M.; de Crécy-Lagard, V.; Diaz, N.; Disz, T.; Edwards, R.; et al. The Subsystems Approach to Genome Annotation and its Use in the Project to Annotate 1000 Genomes. *Nucleic Acids Res.* **2005**, *33*, 5691–5702. [CrossRef]
54. Henning, C.; Gautam, D.; Muriana, P. Identification of Multiple Bacteriocins in *Enterococcus* spp. Using an *Enterococcus*-Specific Bacteriocin PCR Array. *Microorganisms* **2015**, *3*, 1–16. [CrossRef] [PubMed]
55. He, Q.; Hou, Q.; Wang, Y.; Li, J.; Li, W.; Kwok, L.-Y.; Sun, Z.; Zhang, H.; Zhong, Z. Comparative Genomic Analysis of *Enterococcus faecalis*: Insights into Their Environmental Adaptations. *BMC Genom.* **2018**, *19*, 527. [CrossRef] [PubMed]
56. Kohler, V.; Vaishampayan, A.; Grohmann, E. Broad-Host-Range Inc18 Plasmids: Occurrence, Spread and Transfer Mechanisms. *Plasmid* **2018**, *99*, 11–21. [CrossRef] [PubMed]
57. Liu, Y.; Wang, Y.; Schwarz, S.; Li, Y.; Shen, Z.; Zhang, Q.; Wu, C.; Shen, J. Transferable Multiresistance Plasmids Carrying *cfr* in *Enterococcus* spp. from Swine and Farm Environment. *Antimicrob. Agents Chemother.* **2013**, *57*, 42–48. [CrossRef] [PubMed]
58. Devi, S.M.; Halami, P.M. Detection of Mobile Genetic Elements in Pediocin PA-1 Like Producing Lactic Acid Bacteria. *J. Basic Microbiol.* **2013**, *53*, 555–561. [CrossRef] [PubMed]
59. Nicoloff, H.; Bringel, F. IS *Lpl1* Is a Functional IS *30* -Related Insertion Element in *Lactobacillus plantarum* That Is Also Found in Other Lactic Acid Bacteria. *Appl. Environ. Microbiol.* **2003**, *69*, 6032–6040. [CrossRef]
60. Hendrickx, A.P.A.; van Luit-Asbroek, M.; Schapendonk, C.M.E.; van Wamel, W.J.B.; Braat, J.C.; Wijnands, L.M.; Bonten, M.J.M.; Willems, R.J.L. SgrA, a Nidogen-Binding LPXTG Surface Adhesin Implicated in Biofilm Formation, and EcbA, a Collagen Binding MSCRAMM, Are Two Novel Adhesins of Hospital-Acquired *Enterococcus faecium*. *Infect. Immun.* **2009**, *77*, 5097–5106. [CrossRef]
61. Margalho, L.P.; van Schalkwijk, S.; Bachmann, H.; Sant'Ana, A.S. *Enterococcus* spp. in Brazilian artisanal cheeses: Occurrence and Assessment of Phenotypic and Safety Properties of a Large Set of Strains Through the Use of High Throughput Tools Combined with Multivariate Statistics. *Food Control* **2020**, *118*, 107425. [CrossRef]
62. Sun, Z.; Harris, H.M.B.; McCann, A.; Guo, C.; Argimón, S.; Zhang, W.; Yang, X.; Jeffery, I.; Cooney, J.; Kagawa, T.F.; et al. Expanding the Biotechnology Potential of Lactobacilli Through Comparative Genomics Of 213 Strains and Associated Genera. *Nat. Commun.* **2015**, *6*, 8322. [CrossRef]
63. Tarrah, A.; Pakroo, S.; Junior, W.J.F.L.; Guerra, A.F.; Corich, V.; Giacomini, A. Complete Genome Sequence and Carbohydrates-Active EnZymes (CAZymes) Analysis of *Lactobacillus paracasei* DTA72, a Potential Probiotic Strain with Strong Capability to Use Inulin. *Curr. Microbiol.* **2020**, *77*, 2867–2875. [CrossRef]
64. Armenta, S.; Moreno, S.A.; Sánchez-Cuapio, Z.; Sánchez, S.; Rodríguez-Sanoja, R. Advances in Molecular Engineering of Carbohydrate-Binding Modules. *Proteins Struct. Funct. Bioinform.* **2017**, *85*, 1602–1617. [CrossRef] [PubMed]
65. Vaaje-Kolstad, G.; Westereng, B.; Horn, S.J.; Liu, Z.; Zhai, H.; Sørlie, M.; Eijsink, V.G.H. An Oxidative Enzyme Boosting the Enzymatic Conversion of Recalcitrant Polysaccharides. *Science* **2010**, *330*, 219–222. [CrossRef] [PubMed]
66. Wong, E.; Vaaje-Kolstad, G.; Ghosh, A.; Hurtado-Guerrero, R.; Konarev, P.V.; Ibrahim, A.F.M.; Svergun, D.I.; Eijsink, V.G.H.; Chatterjee, N.S.; Van Aalten, D.M.F. The *Vibrio cholerae* Colonization Factor GbpA Possesses a Modular Structure that Governs Binding to Different Host Surfaces. *PLoS Pathog.* **2012**, *8*, e1002373. [CrossRef]
67. Kiousi, D.E.; Efstathiou, C.; Tegopoulos, K.; Mantzourani, I.; Alexopoulos, A.; Plessas, S.; Kolovos, P.; Koffa, M.; Galanis, A. Genomic Insight into *Lacticaseibacillus paracasei* SP5, Reveals Genes and Gene Clusters of Probiotic Interest and Biotechnological Potential. *Front. Microbiol.* **2022**, *13*, 2038. [CrossRef] [PubMed]
68. Chilambi, G.S.; Nordstrom, H.R.; Evans, D.R.; Ferrolino, J.A.; Hayden, R.T.; Marón, G.M.; Vo, A.N.; Gilmore, M.S.; Wolf, J.; Rosch, J.W.; et al. Evolution of Vancomycin-Resistant *Enterococcus faecium* During Colonization and Infection in Immunocompromised Pediatric Patients. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 11703–11714. [CrossRef] [PubMed]
69. Zhang, X.; Vrijenhoek, J.E.P.; Bonten, M.J.M.; Willems, R.J.L.; van Schaik, W. A Genetic Element Present on Megaplasmids Allows *Enterococcus faecium* to Use Raffinose as Carbon Source. *Environ. Microbiol.* **2011**, *13*, 518–528. [CrossRef]
70. Saber, M.M.; Shapiro, B.J. Benchmarking Bacterial Genome-Wide Association Study Methods Using Simulated Genomes and Phenotypes. *Microb. Genom.* **2020**, *6*, e000337. [CrossRef]