

Comparative Genomics of Multidrug-Resistant Enterococcus spp. isolated from Wastewater Treatment Plants

Haley Sanderson

Agriculture and Agri-Food Canada

Rodrigo Ortega-Polo

Agriculture and Agri-Food Canada

Rahat Zaheer

Agriculture and Agri-Food Canada

Noriko Goji

Canadian Food Inspection Agency

Kingsley K. Amoako

Canadian Food Inspection Agency

R. Stephen Brown

Queen's University

Anna Majury

Public Health Ontario

Steven N. Liss

Ryerson University

Tim Mcallister (✉ tim.mcallister@canada.ca)

Agriculture and Agri-Food Canada <https://orcid.org/0000-0002-8266-6513>

Research article

Keywords: Vancomycin resistant Enterococcus (VRE), enterococci, genomics, antimicrobial resistance, mobilome, genome size, wastewater, pangenome, biomarker

Posted Date: December 5th, 2019

DOI: <https://doi.org/10.21203/rs.2.10611/v5>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published on January 24th, 2020. See the published version at <https://doi.org/10.1186/s12866-019-1683-4>.

Abstract

Background Wastewater treatment plants (WWTPs) are considered hotspots for the environmental dissemination of antimicrobial resistance (AMR) determinants. Vancomycin-Resistant Enterococcus (VRE) are candidates for gauging the degree of AMR bacteria in wastewater. Enterococcus faecalis and Enterococcus faecium are recognized indicators of fecal contamination in water. Comparative genomics of enterococci isolated from conventional activated sludge (CAS) and biological aerated filter (BAF) WWTPs was conducted. Results VRE isolates, including E. faecalis (n=24), E. faecium (n=11), E. casseliflavus (n=2) and E. gallinarum (n=2), were selected for sequencing based on WWTP source, species and AMR phenotype. The pangenomes of E. faecium and E. faecalis were both open. The genomic fraction related to the mobilome was positively correlated with genome size in E. faecium ($p < 0.001$) and E. faecalis ($p < 0.001$) and with the number of AMR genes in E. faecium ($p = 0.005$). Genes conferring vancomycin resistance, including van A and van M (E. faecium), van G (E. faecalis), and van C (E. casseliflavus / E. gallinarum), were detected in 20 genomes. The most prominent functional AMR genes were efflux pumps and transporters. A minimum of 16, 6, 5 and 3 virulence genes were detected in E. faecium, E. faecalis, E. casseliflavus and E. gallinarum, respectively. Virulence genes were more common in E. faecalis and E. faecium, than E. casseliflavus and E. gallinarum. A number of mobile genetic elements were shared among species. Functional CRISPR/Cas arrays were detected in 13 E. faecalis genomes, with all but one also containing a prophage. The lack of a functional CRISPR/Cas arrays was associated with multi-drug resistance in E. faecium. Phylogenetic analysis demonstrated differential clustering of isolates based on source but not based on WWTP. Genes related to phage and CRISPR/Cas arrays could potentially serve as environmental biomarkers. Conclusions There was no discernible difference between enterococcal genomes from the CAS and BAF WWTPs. E. faecalis and E. faecium have smaller genomes and harbor more virulence, AMR, and mobile genetic elements than other Enterococcus spp.

Background

Enterococci are ubiquitous in nature and can be found in a variety of environments, including soil, plants, surface water, wastewater, food, and the gastrointestinal tract of animals and humans (Muller et al, 2001; Sanderson et al, 2016). *Enterococcus faecalis* and *Enterococcus faecium*, are associated with a variety of clinical infections of the urinary tract, heart, surgical wounds, bloodstream and neonates (Strateva et al, 2016) and are indicators of fecal contamination (Boehm & Sassoubre, 2014). The ability to treat infections caused by *Enterococcus spp.* is hindered by the development and spread of antimicrobial resistance (AMR) (Arias and Murray, 2008). Resistance to antimicrobials of last resort, such as vancomycin, impairs the ability to control enterococcal infections and is usually accompanied by resistance to other antimicrobials (Klein, 2003; Hammerum, 2012).

Enterococci and antimicrobials are excreted in urine and feces, and in urbanized developed nations, most of this waste is transported to and treated in wastewater treatment plants (WWTPs) prior to discharge into the environment. WWTPs could be considered points of control for the environmental dissemination

of AMR and ideal environments to investigate the epidemiology of AMR from a “One Health” perspective (Ashbolt et al, 2013; Munck et al, 2015; Roca et al, 2015). Within this environment, enterococci can not only exchange genes coding for AMR, but also for heavy metal resistance as well as other genes that increase persistence and survival in other environments (Baker-Austin et al, 2006). This outcome can facilitate the broader dissemination of AMR genes (Ashbolt et al, 2013). Comparative genomics has been applied to identify genes responsible for virulence, AMR, metabolism, secondary metabolite production and gene mobility. Comparative genomics can also be used to compare genes from other functional categories, to predict the ecological fitness of strains, and to discern evolutionary relationships among species.

We previously isolated a number of species of enterococci from two WWTPs with different treatment processes, a conventional activated sludge (CAS) and a biological aerated filter (BAF) system, with *E. faecalis* being the dominant species identified (Sanderson et al, 2019). This work demonstrated changes in AMR phenotypes between wastewater enterococci before and after treatment and between WWTPs. In the current study, we selected 39 wastewater enterococci for sequencing out of 1111 enterococci isolated, including 308 that exhibited vancomycin resistance in broth culture. Isolates were selected so as to be representative of before and after treatment in both WWTPs (Sanderson et al, 2019). We hypothesized that the genomes would not cluster by treatment process but genomes from the BAF system may contain more biofilm-related genes than those from the CAS system. We also proposed that there would be more virulence, AMR, and genetic mobility genes present in *E. faecalis* and *E. faecium* than other *Enterococcus spp.* and that the larger genomes in these clinically relevant species would correlate with the number of mobile genetic elements and genes conferring fitness for survival in a broader range of environments.

Results And Discussion

Sequence Statistics and Pan-genomic Analysis

A summary of sequencing statistics for the 39 *Enterococcus spp.* genomes can be found in Table 1. The genomes ranged from 2.48-3.54 Mbp. The genomes of *E. casseliflavus* and *E. gallinarum* (3.37-3.54 Mbp, 3161-3344 genes) were larger than those of *E. faecalis* (2.69-3.09 Mbp, 2528-3051 genes) and *E. faecium* (2.48-3.02 Mbp, 2374-2992 genes). There was considerable variation in the size of the genomes within a species as a result of differences in the size of the chromosome and the presence/absence of plasmids. The GC content of the genomes ranged from 37.3-37.7%, 37.5-38.1%, and 40.4-42.9% for *E. faecalis*, *E. faecium*, and *E. casseliflavus/E. gallinarum*, respectively.

The range in contigs generated during sequencing was greater in *E. faecium* (21-195 contigs) than in other species (11-68 contigs), likely due to the presence of repetitive and insertion genetic elements complicating assembly (Ricker et al, 2012). Genome sizes were greater for vancomycin and multi-drug resistant strains of *E. faecium* (3.04 Mbp) than for susceptible strains (2.60 Mbp). The genome size of vancomycin-resistant and multi-drug resistant *E. faecalis* was similar to their susceptible counterparts.

The *E. faecalis* pangenome consisted of 5708 genes with a core of 2054 genes (36%), a soft-core of 91 genes (1.6%), a shell genome of 1193 (20.9%) and a cloud genome of 2370 genes (41.5%; Figure 1a). The *E. faecium* pangenome consisted of 3950 genes with a core of 1959 genes (49.6%), a shell of 1186 genes (30%) and a cloud genome of 805 genes (20.4%; Figure 1b). The pan-genome of *Enterococcus spp.* is considered open as it is continually expanding and acquiring new accessory genome elements from other enterococci and bacterial species (Zhong et al, 2017).

Table 1 Genome Characteristics of *Enterococcus spp.* Isolated from Municipal Wastewater Treatment Plants.

Strain	Location	Species	# of Contigs	Size (bp)	%GC	Genes	CDSs	ST*
B72	BAF FE	<i>E. casseliflavus</i>	32	3538396	42.8	3344	3283	NA
B79	BAF FE	<i>E. casseliflavus</i>	49	3527325	42.9	3327	3268	NA
W41	CAS PE	<i>E. faecalis</i>	21	2693209	37.7	2528	2471	116
B139	BAF FE	<i>E. faecalis</i>	30	2720730	37.7	2553	2496	138/501
W314	CAS PE	<i>E. faecalis</i>	11	2721427	37.6	2583	2524	277
C34	CAS PE	<i>E. faecalis</i>	13	2731087	37.6	2615	2556	715
R95	BAF PE	<i>E. faecalis</i>	38	2761310	37.6	2596	2538	674
W350	CAS FE	<i>E. faecalis</i>	28	2789796	37.6	2731	2673	84
R76	BAF FE	<i>E. faecalis</i>	51	2800339	37.6	2690	2627	16
C106	CAS PE	<i>E. faecalis</i>	24	2817683	37.4	2670	2610	16
B48	BAF PE	<i>E. faecalis</i>	29	2822491	37.5	2701	2641	16
B168	BAF PE	<i>E. faecalis</i>	39	2834215	37.5	2725	2667	21
W191	CAS FE	<i>E. faecalis</i>	22	2839739	37.6	2745	2684	207
W460	CAS FE	<i>E. faecalis</i>	27	2848194	37.4	2733	2674	672
H120S2	BAF PE	<i>E. faecalis</i>	21	2853021	37.4	2731	2670	16
C33	CAS PE	<i>E. faecalis</i>	19	2860595	37.3	2722	2662	16
R378	BAF FE	<i>E. faecalis</i>	50	2892126	37.5	2858	2796	326
W75	CAS FE	<i>E. faecalis</i>	59	2901424	37.5	2838	2776	209
B6	BAF FE	<i>E. faecalis</i>	48	2906126	37.5	2796	2738	26
R395	BAF PE	<i>E. faecalis</i>	35	2951239	37.5	2844	2785	40
W195	CAS FE	<i>E. faecalis</i>	30	2970793	37.5	2865	2806	40
H114S2	BAF PE	<i>E. faecalis</i>	35	2979979	37.4	2881	2822	40
C379	CAS FE	<i>E. faecalis</i>	42	2988783	37.5	2902	2844	40
R61	BAF FE	<i>E. faecalis</i>	37	3004659	37.3	2969	2907	16
B150	BAF FE	<i>E. faecalis</i>	41	3012117	37.3	2955	2893	16
W37	CAS PE	<i>E. faecalis</i>	68	3088982	37.3	3051	2990	768
C329	CAS FE	<i>E. faecium</i>	63	2480628	38.1	2374	2305	40
B466	BAF FE	<i>E. faecium</i>	71	2553406	38.1	2512	2443	672

C567	CAS FE	<i>E. faecium</i>	21	2778016	37.9	2714	2644	1216
C12d	CAS PE	<i>E. faecium</i>	169	2879332	37.7	2823	2755	18
H101S2	BAF PE	<i>E. faecium</i>	177	2884826	37.8	2836	2770	18
H123S2	BAF PE	<i>E. faecium</i>	170	2912775	37.7	2871	2803	18
H53S1	CAS PE	<i>E. faecium</i>	165	2959005	37.6	2915	2847	18
F11j	CAS PE	<i>E. faecium</i>	181	2964368	37.6	2916	2848	18
R407	BAF PE	<i>E. faecium</i>	176	3005175	37.5	2955	2887	18
B492	BAF FE	<i>E. faecium</i>	189	3008305	37.5	2961	2893	18
R337	BAF PE	<i>E. faecium</i>	195	3023784	37.5	2992	2924	18
W17	CAS FE	<i>E. gallinarum</i>	14	3367806	40.5	3161	3102	NA
G12s	CAS PE	<i>E. gallinarum</i>	30	3442529	40.4	3303	3244	NA

BAF: biological aerated filter, CAS: conventional activated sludge, PE: primary effluent, FE: final effluent, % GC: guanine-cytosine content, CDS: protein coding sequence, ST: sequence type, NA: not available

Multi-locus Sequence Typing

In the current study, four sequence types (STs) for *E. faecium* and 15 STs for *E. faecalis* were identified (Table 1). Eight *E. faecium* genomes belonged to ST18, part of the clonal complex 17 (CC-17) associated with clinical infections and in treated and untreated wastewater (Caplin et al, 2008). This suggests that the majority of *E. faecium* sequenced from wastewater originated from humans. Out of the *E. faecalis* STs identified in this study, ST16 (n=7) and ST40 (n=4) were the most common and have previously been associated with high level gentamicin resistance in clinical isolates and in isolates from pigs (Ruiz-Garbajosa et al, 2006; Hammerum, 2012). However, only 5 of the isolates in this study (4 *E. faecalis* and 1 *E. faecium*) exhibited high level gentamicin resistance. This may differ between studies because of geographical location and because gentamicin resistance is transferable and may not be present in all ST16 and ST40 *E. faecalis* isolates.

Phylogeny

Genomes did not cluster based on WWTP, but all species formed separate monophylogenetic groups (Figure 2). This finding supports our previous results of speciation of enterococci based on the *groESL* locus (Zaheer et al, 2012; Sanderson et al, 2019). The majority of wastewater *E. faecalis* isolates were more closely related to livestock and food-derived *E. faecalis* genomes, while seven wastewater strains (B139, B168, C34, W37, W75, W191, and W314) clustered with strains isolated from human infections (Figure 3). The diversity of wastewater strains maybe a reflection of their origin from clinical, companion

animal or agricultural sources. None of the *E. faecalis* wastewater, human, and agriculture (and food-derived) isolates clustered together by source, suggesting that agricultural and human clinical strains are phylogenetically distinct. Vancomycin-resistant *E. faecalis* isolates also did not cluster as they belonged to different STs, unlike vancomycin-resistant *E. faecium*, which did cluster as all isolates belonged to CC-17 (Figure 4). There was more genetic diversity in vancomycin-resistant *E. faecalis* than *E. faecium*. For *E. faecium*, wastewater strains clustered separately from most clinical strains (Figure 4). The bovine strain *E. faecium* F1129F clustered with human clinical strains, whereas the other bovine strain, *E. faecium* F1213D did not. Three wastewater isolates (*E. faecium* C567, *E. faecium* B466, and *E. faecium* C329) were more closely related to *E. faecium* F1213D (bovine) and *E. faecium* NRRL B-2354 (food) than to clinical isolates. The distinct clustering between clinical and wastewater strains of *E. faecium* may be due to the large accessory genome and characterization of the genes may provide insight into the mechanisms whereby enterococci adapt to specific environments.

Clusters of Orthologous Groups (COGs): Functional Categories and Genome Size

Clusters of Orthologous Groups (COGs) are broad functional categories used to assign proteins to their specific function (Tatusov et al, 2001). Functional categorization of proteins into different COGs revealed variation profiles among *Enterococcus spp.*, but little difference among strains within species, with the exception of the mobilome and genes associated with energy production and conversion (Additional File 1, sheet 6). A disproportionate increase in genes associated with energy conversion, regulatory function, transport and secondary metabolism has been noted with expansion in genome size in previous comparative bacterial genomic studies (Stover et al, 2000; Bentley et al, 2002; Konstantinidis and Tiedje, 2004). We assessed which functional categories of genes were disproportionately represented in the isolates collected from the WWTPs with expanded genomes so as to determine if some of these COGs could be increasing the fitness of multi-drug resistant enterococci. This could ultimately increase the risk of infection with these strains and the transfer of virulence and AMR determinants to other bacteria.

Given the variation in genome size between and within species, the relationships between genome size and the number of genes associated with specific functional categories was determined (Figure 5; Additional File, Sheet 6). There were more COGs assigned to carbohydrate transport and metabolism, transcription, cell motility, secondary metabolite biosynthesis, transport, catabolism and signal transduction mechanisms in *E. casseliflavus* and *E. gallinarum* compared to enterococci more frequently associated with clinical infections. These functional categories could allow for higher fitness in aquatic environments where more diverse substrates are typically available at much lower concentrations than in the digestive tract. The increase in cell motility related genes may also enable these species to undertake chemotaxis in aquatic environments where nutrients may be scarce (Roszak and Colwell, 1987). Compared to *E. faecalis* and *E. Faecium*, these genomes also contained more genes encoding for secondary metabolites including antimicrobial agents, compared to *E. faecalis* and *E. faecium* which although not required for growth can confer competitiveness in diverse environments (Kell et al, 1995). *E. casseliflavus* and *E. gallinarum* are known to be more environmentally fit than *E. faecalis* and *E. faecium* as a result of a variety of mechanisms. For instance, the yellow pigment of *E. casseliflavus* can protect

the species from photo-inactivation in aquatic environments (Maraccini et al, 2012). However, *E. faecium* and *E. faecalis* are still the predominant species in wastewater, likely due to the continuous input of fecal waste which sustains their survival.

When all of the wastewater *Enterococcus* genomes were pooled, there was a strong negative correlation ($p < 0.001$) between genome size and nucleotide transport and metabolism, lipid metabolism and translation, ribosomal structure and biogenesis and a strong positive correlation ($p < 0.001$) between genome size and cell motility (Figure 5a; Additional File 1, sheet 6). The total number of genes related to cell motility, signal transduction, and carbohydrate transport and metabolism were positively correlated ($p < 0.001$) with genome size. This is reflective of the greater genome size of environmental species compared to *E. faecium* and *E. faecalis*. The total number of genes related to cell division and chromosome partitioning, cell envelope biogenesis, outer membrane and post translational modification, protein turnover, and transcription were negatively correlated ($p < 0.001$) with genome size.

The species-specific patterns in genomic proportions for each functional category differ from the pooled genomes for the genus. In both *E. faecalis* and *E. faecium*, a larger genome was strongly correlated with the mobilome ($p < 0.001$) (Figure 5b and 5c), a functional category not included in the analysis of Konstantinidis and Tiedje (2004). In contrast, the mobilome was not correlated with genome size in the pooled *Enterococcus* genomes. This would suggest that the mobilome is a significant factor in the evolution of *E. faecalis* and *E. faecium* from wastewater and may contribute to the genomic expansion and diversity in wastewater *Enterococcus* species. There was also a positive correlation ($p = 0.005$) between the number of unique AMR genes and genome size of *E. faecium*, suggesting the acquisition of AMR genes occurs through horizontal gene transfer. For example, *E. faecium* R337 has a genome of 3.02 kbp, 58 genes associated with the mobilome and 23 AMR genes; while *E. faecium* C329 had a genome of 2.48kbp and 15 genes associated with the mobilome and 3 AMR genes.

The total number of genes related to cell motility ($p < 0.001$), DNA replication, recombination, and repair ($p < 0.001$), extracellular structures ($p < 0.001$), and mobilome ($p < 0.001$) was positively correlated with genome size in *E. faecium*. The number of AMR genes also showed a positive correlation ($p = 0.002$) with the amount of genes related to the mobilome in this species (Figure 5c). The eight *E. faecium* genomes belonged to the same sequence type (CC-17), while *E. faecalis* genomes were more diverse suggesting that *E. faecium* isolates may originate from and be more specifically adapted to clinical environments.

Antimicrobial Resistance Genes

Vancomycin-resistant enterococci have been known to exhibit resistance to a number of antimicrobials (Klein, 2003; Varela et al, 2013). Enterococci are also intrinsically resistant to beta-lactams, aminoglycosides and streptogramins and can acquire antimicrobial resistance through horizontal gene transfer (Klein, 2003; Varela et al, 2013; Miller et al, 2014). In this study, we screened 39 multi-antimicrobial resistant enterococci genomes against the CARD database for antimicrobial resistance genes (ARGs) (Table 3; Additional Material 1, Sheet 8) and ten genes (*eatAv*, *emeA*, *lsaA*, *efrA*, *efrB*, *tetL*,

efmA, *msrC*, *ermY*, and *IsaE*) associated with multi-drug efflux pumps and other transporters were detected. These efflux proteins may confer intermediate resistance to a variety of antimicrobials.

Genes conferring glycopeptide (vancomycin and teicoplanin) resistance were detected in 20 of the genomes. In *E. faecium* and *E. faecalis*, resistance was conferred by *vanA* and *vanM* in *E. faecium* or *vanG* in *E. faecalis*. Vancomycin resistance was mediated by *vanC*, and this was the only ARG detected, in *E. casseliflavus* and *E. gallinarum*. There are a variety of ARGs that confer vancomycin resistance, with *vanA*, *vanB* and *vanC* being the most common in wastewater enterococci. The most common determinant for teicoplanin resistance is *vanZ*, which can be integrated into the *van* operon, although it is absent in the *vanB* operon, and confers resistance to both vancomycin and teicoplanin (Faron et al, 2016). As a result, teicoplanin resistance is commonly associated with vancomycin resistance. Although rarely, teicoplanin resistance without vancomycin resistance is likely due to changes in the promoter of *van* operon or due to the presence of a different resistance mechanism (Chang et al, 2013; Grove, 2013; Loong et al, 2016).

ErmB confers resistance to macrolide-lincosamide-streptogramin B (MLS_B) antimicrobials and was found in *E. faecium* (n=7) and *E. faecalis* (n=4) among our isolates. Other *erm* genes (*ermC*, *ermG*, *ermT*, *ermY*) were detected in one *E. faecium* genome. *msrC*, which codes for a macrolide efflux pump, was only detected in *E. faecium* (n=11). Resistance to erythromycin and other macrolides can arise as a result of mutations in the 23S rRNA gene or by efflux pumps (Miller et al 2014). Macrolides are used extensively in both human and animal health. Blanch et al (2003) observed that most wastewater isolates with high-level vancomycin resistance were also resistant to erythromycin, suggesting that erythromycin resistance may favour the persistence of VRE in the environment. The modification of the 23S rRNA target by methylase genes, like *ermB*, can also confer resistance to streptogramins (Miller et al 2014). The most common macrolide resistance gene detected in enterococcal genomes was *ermB* (n=15).

Enterococci exhibit intrinsic resistance to low concentrations of aminoglycosides as a result of the presence of *aac(6')-II*. Thirteen of the enterococci isolates were resistant to high concentrations of gentamicin and streptomycin. These aminoglycosides are clinically-important as they are not inactivated by *aac(6')-II*; and *E. faecium* are typically sensitive to these antimicrobials (Miller et al, 2014). In our study, cross-resistance to levofloxacin and the aminoglycosides (gentamicin and streptomycin) occurred in 5 isolates with 3 additional isolates exhibiting intermediate resistance to one or more of these antimicrobials. Aside from cross-resistance to other classes of antimicrobials, like fluoroquinolones, resistance to these aminoglycosides is likely acquired. In our study, additional aminoglycoside genes (*ant(9')-Ia*, *aad(6')*, *aph(3')-IIIa*, *SAT-4*, *ant(6')-Ia*, and *aac(6')-Ie-aph(2'')-Ia*) were detected in the genomes of up to 5 *E. faecalis* and 7 *E. faecium* aminoglycoside resistant isolates. Others have shown that these resistance genes are frequently encoded on plasmids and transposons (Miller et al, 2014). Streptomycin resistance either involves the inhibition of the drug at the ribosomal level or enzyme inactivation by an acquired streptomycin adenylyltransferase (Miller et al, 2014). Gentamicin resistance arises as the result of the acquisition of *aac(6')-Ie-aph(2'')-Ia*, which was detected in 7 genomes (2 *E. faecalis* and 5 *E. faecium*) and confers resistance to all aminoglycosides except streptomycin (Miller et al, 2014). The prevalence of

streptomycin resistance versus gentamicin resistance differed between species, with streptomycin resistance being more common in *E. faecium* and gentamicin resistance more common in *E. faecalis*.

Genes encoding tetracycline resistance were detected in 26 of the genomes, including *E. faecium* and *E. faecalis*. *Tet(L)* encodes an efflux protein and *tet(M)* and *tet(S)* encode for ribosomal protection proteins. Disk susceptibility testing revealed that these isolates were resistant to doxycycline, whilst those containing *tet(L)* were susceptible, suggesting specificity for the *tet(L)* efflux protein. In general, bacteria that are resistant to doxycycline are also resistant to tetracycline and oxytetracycline (Roberts et al, 2002; Holzel et al, 2010). Tetracycline resistance can be due to efflux pumps or ribosomal protection mechanisms, which can be chromosomal and/or plasmid-borne. Co-selection of tetracycline and macrolide resistance in environmental enterococci may occur (Martins da Costa et al, 2006ab). In this study, determinants for macrolide and tetracycline were detected together in 16 of the enterococcal genomes.

Genes associated with resistance to antimicrobials not included in the disc susceptibility panel were also detected. A gene associated with chloramphenicol resistance, *cat*, was detected in two *E. faecalis* genomes. Genes associated with diaminopyrimidine resistance (*dfpE*, *dfpF*, and *dfpG*) were also detected in *E. faecium* and *E. faecalis*. Two *E. faecalis* genomes also had genes associated with lincosamide resistance (*InuB* and *InuG*).

Table 2 Phenotypic Antimicrobial Resistance Profiles of sequenced *Enterococcus spp* isolates sequenced.

Isolate	AMR Phenotype*
<i>E. casseliflavus</i> B72	ERY ^I , Q-DR
<i>E. casseliflavus</i> B79	VAN ^I , ERY ^I , Q-D ^I
<i>E. faecalis</i> W460	ERY ^I , Q-DR
<i>E. faecalis</i> W350	DOX ^R , ERY ^I , LZD ^I
<i>E. faecalis</i> W191	DOX ^R , Q-DR
<i>E. faecalis</i> H120S2	ERY ^R , STR ^I
<i>E. faecalis</i> H114S2	VAN ^R , TEC ^R , AMP ^R , LVX ^R
<i>E. faecalis</i> C34	VAN ^R , TEC ^R , ERY ^I , LZD ^I , Q-DR
<i>E. faecalis</i> C33	TEC ^R , DOX ^R , ERY ^R , LZD ^I , LVX ^R , Q-DR
<i>E. faecalis</i> B168	DOX ^I , ERY ^I , LZD ^I , Q-DR
<i>E. faecalis</i> B150	TEC ^R , DOX ^I , ERY ^R , LZD ^R , Q-DR
<i>E. faecalis</i> B6	DOX ^R , ERY ^I , Q-DR
<i>E. faecalis</i> B48	DOX ^R , ERY ^R , LZD ^R , Q-DR
<i>E. faecalis</i> B139	VAN ^I , LZD ^I , Q-DR
<i>E. faecalis</i> C106	TEC ^R , DOX ^R , ERY ^R , GEN ^R , Q-DR
<i>E. faecalis</i> C379	DOX ^R , ERY ^I , Q-DR
<i>E. faecalis</i> R61	VAN ^I , TEC ^R , DOX ^R , ERY ^R , GEN ^I , LZD ^I , Q-DR
<i>E. faecalis</i> R76	DOX ^R , ERY ^R , GEN ^R , Q-DR, STR ^R
<i>E. faecalis</i> R95	DOX ^I , ERY ^R , GEN ^I , LZD ^I , Q-DR, STR ^R
<i>E. faecalis</i> R378	ERY ^I , Q-D ^I
<i>E. faecalis</i> R395	VAN ^I , TEC ^R , DOX ^R , ERY ^I , Q-DR
<i>E. faecalis</i> W37	DOX ^I , ERY ^R , LVX ^R , Q-D ^I , STR ^R
<i>E. faecalis</i> W41	VAN ^I , TEC ^R , DOX ^R , ERY ^I , Q-DR
<i>E. faecalis</i> W75	VAN ^I , Q-D ^I
<i>E. faecalis</i> W195	DOX ^I , ERY ^I , Q-DR
<i>E. faecalis</i> W314	TEC ^R , ERY ^I , LZD ^R , Q-DR
<i>E. faecium</i> R407	VAN ^R , TEC ^R , AMP ^R , ERY ^R , GEN ^R , LVX ^R , STR ^R
<i>E. faecium</i> R337	VAN ^R , TEC ^R , AMP ^R , ERY ^R , NIT ^I , LVX ^R , STR ^R
<i>E. faecium</i> H53S1	VAN ^R , TEC ^R , AMP ^R , ERY ^R , LVX ^R , STR ^I
<i>E. faecium</i> F11J	VAN ^R , TEC ^R , AMP ^R , ERY ^R , LVX ^R , STR ^R
<i>E. faecium</i> C329	TEC ^R , ERY ^I , NIT ^R , LZD ^I , LVX ^R
<i>E. faecium</i> B492	VAN ^R , TEC ^R , AMP ^R , ERY ^R , LVX ^R , STR ^R
<i>E. faecium</i> B466	AMP ^R , DOX ^R , ERY ^R , NIT ^I , LVX ^I
<i>E. faecium</i> C12D	VAN ^R , TEC ^R , AMP ^R , ERY ^I , LVX ^R
<i>E. faecium</i> C567	ERY ^R , NIT ^I , LZD ^I
<i>E. faecium</i> H101S2	VAN ^R , TEC ^R , AMP ^R , ERY ^R , LVX ^R , STR ^I
<i>E. faecium</i> H123S2	VAN ^R , TEC ^R , AMP ^R , ERY ^R , LVX ^R , STR ^I
<i>E. gallinarum</i> W17	VAN ^I
<i>E. gallinarum</i> G12S	VAN ^I

*AMR phenotypic profiles using R for resistant to the antimicrobial and I for intermediately resistant to the antimicrobial. Antimicrobials used for disc susceptibility testing were vancomycin (VAN), teicoplanin (TEC), ampicillin (AMP), doxycycline (DOX), erythromycin (ERY), levofloxacin (LVX), linezolid (LZD),

nitrofurantoin (NIT), gentamicin (GEN), streptomycin (STR), quinupristin/dalfopristin (Q-D), and tigecycline (TGC).

Virulence Genes

The number of shared virulence genes among genomes of the same species were 16, 6, 5 and 3 for *E. faecium*, *E. casseliflavus*, *E. faecalis*, and *E. gallinarum*, respectively (Additional Material 1, Sheet 9-11). All of the *E. faecium* isolates contained genes related to adhesion to surfaces (*tuf*, *aga*, *efaA*, and *sgrA*), cell wall biosynthesis (phosphatase cytidylyltransferase, *uppS*), cellular defense (*lisR*), biofilm formation and surface proteins (*acm*, *esp*, *scm* and type A and B pili). Other functions including bile salt degradation (*bsh*), proteases (*tip/ropA*), biofilm formation (*bopD*), enolase (*eno*), and antiphagocytosis and capsule formation (*rfaA-1*) were also identified. All of the *E. faecalis* genomes contained genes for cell adhesion (*tuf*), carbohydrate metabolism (*hyl*), endocarditic and biofilm association (ebp) pili (*ebpA*), Type III secretion proteins (*bopD*) and fibrinogen-binding proteins (*fss1*). All of the *E. casseliflavus* genomes contained the same five virulence genes with functions of: capsule biosynthesis (*capE*), enolase (*eno*), leucine aminopeptidase (*lap*), heat shock protein (*hsp60*), and protein modification (*lplA1*). All of the *E. gallinarum* genomes had an enolase (*eno*), a flagellar biosynthesis protein (*flhA*) and a bile salt hydrolase (*bsh*). One of the *E. gallinarum* genomes also contained genes related to capsule proteins and the other, originating from effluent possessed in addition to the CAS system, 2 genes associated with metal transporter (*ssaB* and *psaA*).

The virulence genes detected have additional functions for improved environmental fitness. For instance, the majority of the virulence genes detected in the genomes from this study were also associated with biofilm formation or adherence to surfaces (i.e., *ace*, *acm*, *agg*, *bop*, *ccf*, *cob*, *cpd*, *ebpABC*, *ecbA*, *efaA*, *esp*, *fsrABC*, *gelE*, *pil*, *scm*, *sgrA*, *sprE*, and *srt*). These genes are ubiquitous as they likely play a role in the fitness of enterococci in both the human digestive tract and WWTPs. A number of capsule protein genes were also common among the genomes and not only confer resistance to phagocytosis in humans and animals (Palmer et al, 2012; Qin et al, 2012), but also to predation by amoeba and bacteriophage in aquatic environments (van Tyne and Gilmore, 2014; Rendueles et al, 2017). Hyaluronidase (*hyl*) genes were detected in all of the *E. faecalis* genomes and have been associated with increased vancomycin resistance and virulence in mouse peritonitis models (Qin et al, 2012).

Mobile Gene Elements

Mobile genetic elements (MGEs) play an important role in horizontal gene transfer and the spread of AMR among isolates in the environment, humans and animal hosts. MGEs include plasmids, transposable elements, prophages and various genomic islands such as integrative conjugative elements (ICE) (van Elsas and Bailey, 2002). ICE and transposons present in the genomes were identified and described using the ICEberg database (Table 4; Additional File 1, sheet 17). The transposon, Tn917 is widely distributed in

enterococci (Shaw and Clewell, 1985) and was identified in 8 of the sequenced *E. faecalis* genomes. All of these strains exhibited erythromycin resistance and *erm*(B) was found to be associated with Tn1545 and Tn917 (Chen et al, 2010). One transposon, Tn6098 was present in all of the genomes and possess genes associated with α -galactoside metabolism. A multidrug resistance transposon, Tn5385, was also found in all of the *E. faecalis* genomes, with these isolates exhibiting erythromycin and doxycycline resistance as this transposon commonly carries these resistance genes (Rice and Carias, 1998). Other Tn5801 and Tn6013-like ICE elements of unknown function were also present in all *E. faecium* isolates, in addition to a cadmium and arsenic resistance ICE, ICESde3396. All of the *E. gallinarum* and *E. casseliflavus* had Tn916-type transposons (Tn6079, Tn6087 and Tn6084, respectively). These transposons can carry genes coding for tetracycline, minocycline and erythromycin resistance (Rice, 1998; Roberts and Mullany, 2011). While these transposons were detected in *E. casseliflavus* and *E. gallinarum*, these isolates did not exhibit erythromycin resistance and no associated AMR genes were detected in their genomes. Seven out of the unique 27 ICE were present in genomes of more than one *Enterococcus* species.

Table 3 Integrative conjugative elements (ICE) and transposons identified in the wastewater *Enterococcus* spp. genomes (n = 39)

Species	Common ICE	Function	Other Notable ICE	Function
<i>E. casseliflavus</i>	Tn6098 Tn6084 Tn6000(EfcTn1)	α -galactoside metabolism Tetracycline resistance Tetracycline resistance	No other ICE detected	
<i>E. faecalis</i>	Tn6098 Tn5385	α -galactoside metabolism Erythromycin, gentamicin, streptomycin, tetracycline, penicillin/ β -lactam, mercury resistance	Tn917 Tn2008 Tn1545 ICESp23FST81 Tn6009 ICESde3396 Tn5301 Tn5276 ICESt1	Tetracycline resistance Chloramphenicol, erythromycin, streptomycin, kanamycin resistance Sulfamethoxazole, trimethoprim, chloramphenicol, erythromycin, streptomycin resistance Tetracycline, chloramphenicol resistance, toxin-antitoxin system Mercury resistance Kanamycin, arsenic and cadmium resistance Nisin biosynthesis Nisin biosynthesis, sucrose fermentation Type II restriction modification system
<i>E. faecium</i>	Tn6098 Tn5801 Tn6084 Tn6000(EfcTn1) ICESauTW20-2 ICESauT0131-2 ICESauJKD6008-2 ICESpsED99-1 ICESauMu3-1	α -galactoside metabolism Tetracycline resistance Tetracycline resistance Tetracycline resistance Unknown Unknown Unknown Unknown Unknown	Tn2008 Tn1545 ICESde3396 ICESt1 ICEAusCo10a-1 10750-RD.2	See above See above See above See above Toxin-antitoxin system Type II restriction modification, spectinomycin, erythromycin resistance
<i>E. gallinarum</i>	Tn6098 Tn6079 Tn6087 ICESde3396	α -galactoside metabolism Tetracycline, erythromycin resistance Tetracycline, antiseptic and antimicrobial resistance (unspecified) Kanamycin, arsenic and cadmium resistance	Tn2008 Tn1545	See above See above

CRISPR-Cas Arrays and Bacteriophage

Type II CRISPR-Cas systems are typically described in enterococci and were detected in 13 *E. faecalis* genomes (Figure 6). Multiple CRISPR arrays can often be detected in bacterial genomes, but not all arrays are accompanied by Cas genes. Orphan CRISPR arrays (without Cas genes) were identified in 27 of the

genomes (Figure 6). Comparison of CRISPR arrays flanked by Cas genes revealed unique arrays among *Enterococcus* species, but some arrays were shared among strains of the same species. Arrays identified in the sequenced *Enterococcus* genomes contained 4 to 20 direct repeat sequences associated with functional CRISPR arrays. An additional 72 unique spacers associated with orphan CRISPR arrays were identified in this study. Eleven *E. faecalis* and 10 *E. faecium* genomes lacked CRISPR-Cas systems. Any genomes lacking functional arrays exhibited resistance to 4 3 or more antimicrobial agents. The absence of CRISPR/Cas systems may compromise genome defence, increasing the likelihood of acquisition of AMR determinants from bacteriophage and plasmids (Palmer and Gilmore, 2010).

When a phage infects a bacterium, it incorporates spacers into the array present in the bacterial chromosome and occasionally plasmids. The spacers are expressed as CRISPR RNAs (crRNAs) and provide a surveillance mechanism for descendant cells and guide the CRISPR/Cas system to enable cleavage of the protospacer sequence in the phage genome. The cleaved phage genomes are then cannibalized and can no longer support productive phage infection (Barrangou and Horvath, 2009; Tao et al, 2018). CRISPR-Cas systems impact the evolution of both bacteria and phage populations. Transduction dependent horizontal gene transfer is a key driver of bacterial evolution and rapid viral evolution to evade CRISPR-Cas systems (Tao et al, 2018). CRISPR/Cas arrays can also provide a record of previous and continued interaction between particular bacteria and phage (Barrangou and Horvath, 2009; Shmakov et al, 2017).

Functional CRISPR arrays and intact prophage were identified in 10 *E. faecalis* genomes, but the combination was not seen in the other 29 genome sequenced in this study. Bacteriophage can develop anti-CRISPR systems to promote their integration into the bacterial genome (Bondy-Denomy et al, 2012). Some of the spacer regions identified in CRISPR arrays were 100% identical to incomplete prophage sequences, but these genomes still contained at least one prophage. These spacers may still limit the type of phage that can integrate into the genome.

Bacteriophage-mediated transduction of AMR has been demonstrated in enterococci and potential virulence determinants have been identified in phage associated with *E. faecalis*. Phages found in the genomes were members of the Siphoviridae and Myoviridae (Additional File 1, Sheet 12). Other prophages in *Enterococcus spp.* belonging to Podoviridae, Inoviridae, Leviridae, Guttaviridae and Fuselloviridae have also been described (Mazaheri Nezhad Fard et al, 2010; Duerkop et al, 2014). Thirty-four of the 39 genomes contained at least one putative phage ranging in size from 19.2kb to 70.6kb. A total of 55 unique intact prophages were identified across 34 sequenced genomes. *E. faecium* and *E. faecalis* contained up to 3 intact prophages, whereas *E. casseliflavus* and *E. gallinarum* contained 1 or 2 intact prophages. Prophages from the Siphoviridae family were the most prevalent across all species and are also commonly identified in lactic acid bacteria (van Schaik et al, 2010).

Secondary Metabolites

Bacteriocins are ribosomally synthesized antimicrobial peptides produced by Gram-positive and Gram-negative bacteria that have antimicrobial activity against closely related bacteria. They could provide a

competitive advantage to the survival of bacteria in ecological niches that exhibit poor nutrient concentrations, heat and extreme pH (Yang et al, 2014). Bacteriocins were identified in 8 *E. faecalis* and 9 *E. faecium* genomes in addition to 1 *E. gallinarum* genome (Additional File 1, Sheet 18). Enterocin A was identified in nine *E. faecium* genomes. Lantipeptides are also a growing class of bacteriocins with a large diversity of activity, structure, and biosynthetic machinery. Lantipeptides have multiple uses including as a limited class of antimicrobials (Knerr and van der Donk, 2012). Lantipeptides were identified in 3 *E. faecalis* genomes as cytolysins, which have both haemolytic and bacteriolytic activities (Booth et al, 1996). Lasso peptides are also ribosomally synthesized and post-translationally modified and can have antimicrobial, enzyme inhibitory, and receptor antagonistic activities (Hegemann et al, 2015). Lasso peptides were identified in 6 *E. faecalis* genomes.

Terpenes were detected in all *E. casseliflavus* and *E. gallinarum*, but not in *E. faecalis* and *E. faecium* genomes. These secondary metabolites are most often associated with plants and fungi, and have been described in prokaryotes in only a few instances, including *Enterococcus spp* (Beukers et al, 2017). Terpenes can have a variety of functions including as antimicrobials, hormones, pigments, and flavor or odour constituents (Paduch et al, 2007), but their role in *Enterococcus spp* is unclear. Aryl polyene was detected in one *E. faecalis* (C34) genome. Aryl polyene biosynthetic clusters produce a pigment that protects the organism from reactive oxygen species (Schoner et al, 2016).

Biomarker Search

The small number of genomes limited the identification of biomarkers, particularly for searches within the same species isolated from different sources (Additional Material 1, Sheet 19). These biomarkers are genes or gene fragments only present in one group of genomes and not others making them possible identifiers of the origin of collected isolates. The majority of searches have identified biomarkers with scores below a correlation cut-off of 0.95). However, in our study, *E. faecalis* from wastewater that clustered with agricultural and animal sources revealed a biomarker associated with CRISPR-associated genes that differentiated (score=0.8043) these isolates from *E. faecalis* from wastewater that clustered with human sources. A comparison of *E. faecium* from clinical (inclusion) and wastewater (exclusion) sources yielded 7 biomarkers with scores greater than 0.80. These were associated with phage (n=6) and hypothetical proteins (n=1). Weigald et al (2014) conducted a similar search with watershed and enteric enterococcal genomes and found shared phenotype and phylogeny between the two groups, but also identified several biomarkers for both sources. These biomarkers encoded accessory nutrient utilization pathways, including a nickel uptake operon and sugar utilization pathways including xylose were overrepresented in enteric genomes (Weigald et al, 2014). A search for potential biomarkers that distinguished among species in this study resulted in 98 signatures for *E. faecalis*, 130 signatures for *E. faecium*, and 3 signatures *E. casseliflavus* and *E. gallinarum*. These signatures include genes related to various types of nucleotide and carbohydrate metabolism, and other genes that can improve environmental fitness, including a variety of transporters and housekeeping genes related to DNA replication, transcription and translation.

Conclusions

In this study, enterococci did not cluster phylogenetically based on point of isolation during waste water treatment or on the type of WWTPs. Despite being the dominant species in wastewater, *E. faecalis* and *E. faecium* have smaller genomes and may be less equipped to survive outside of their niche environment than *E. casseliflavus* and *E. gallinarum*. They do, however, harbor more virulence, AMR, and mobile genetic elements than other *Enterococcus spp.* A larger genome size in *E. faecalis* and *E. faecium* was positively correlated with an expansion in the mobilome. In *E. faecium*, there was a positive correlation between the number of AMR genes and the mobilome. So, while the larger genome size of *E. casseliflavus* and *E. gallinarum* is accompanied by more genes related to metabolism and secondary functions, possibly increasing their fitness in the environmental, this was not the case for *E. faecium* and *E. faecalis*. The study suggests that the key to understanding the impact of WWTPs on AMR dissemination is likely understanding the mobilome and discerning linkages between enterococci in wastewater and other environmental and clinical sources.

Methods

Isolate Selection

Thirty-nine *Enterococcus spp.*, including *E. faecalis* (n=24), *E. faecium* (n=11), *E. casseliflavus* (n=2) and *E. gallinarum* (n=2), isolated from wastewater were selected for whole genome sequencing. These were selected from a collection of 308 isolates collected from the primary and final effluents of two WWTPs in Kingston, Ontario, Canada, a BAF and a CAS system between 2014 and 2016. Isolates were speciated and subsequently underwent disc susceptibility testing for a panel of 12 antimicrobial agents. Nine to ten *Enterococcus* isolates were chosen from each of the primary and final effluent of the two WWTPs to represent the most prominent species isolated from the samples and the most prominent unique antimicrobial resistance phenotypic profiles. While all of these isolates grew in Todd-Hewitt broth supplemented with vancomycin (≥ 4 mg/L), not all met the requirements for vancomycin resistance using disc susceptibility testing following CLSI and EUCAST guidelines. This procedure used reference strains *E. faecium* ATCC 700221 (MIC ≥ 32 mg/L), *E. faecalis* ATCC 51299 (MIC ≥ 4 mg/L) and *E. faecalis* ATCC 29212 (susceptible). Disc susceptibility testing was conducted using CLSI and EUCAST guidelines and the reference strains *Staphylococcus aureus* ATCC 25923 and *Enterococcus faecalis* ATCC 29212. The final isolates selected included 21 vancomycin-susceptible, multi-drug resistant enterococci and 18 enterococci with either intermediate resistance or resistance to vancomycin based on disc susceptibility testing. The AMR phenotypic profiles of the selected isolates are available in Table 3.

DNA Extraction and Sequencing

Enterococcus spp. were grown on Brain Heart Infusion (BHI) agar (Dalylnn Biologicals, Calgary, AB) overnight at 37°C. Colonies from a freshly grown culture plate were suspended in TE buffer to achieve an OD₆₀₀ of 2 in order to harvest 2×10^9 cells, and 1 mL was transferred to a microcentrifuge tube and

centrifuged for 2 min at 14000 x g. Genomic DNA was extracted using a modified DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) with the addition of an enzymatic lysis step. Bacterial cells were incubated at 37°C with shaking (150 rpm) in lysis buffer consisting of 20mM Tris-Cl (pH 8.0), 2 mM sodium EDTA, 1.2% Triton X-100 and 40 mg/mL lysozyme (Sigma Aldrich Canada, Oakville, ON). Proteinase K and 5 µL of 100 mg/mL RNase A were added (Qiagen, Hilden, Germany) subsequently, and incubated at room temperature for 10 min before proceeding to the next step. The quality of the genomic DNA was determined using a Nanodrop One UV-Vis Spectrophotometer (Thermo Scientific, Burlington, ON) and Qubit fluorometer (Thermo Scientific). Genomic library construction was performed using the Illumina Nextera XT DNA sample preparation kit (Illumina Inc., San Diego, CA) following the manufacturer's instructions. The library was sequenced on an Illumina MiSeq platform (Illumina, Inc.). FASTA data was filtered for quality and high-quality reads were *de novo* assembled using SPAdes genome assembler 3.6.0 (Bankevich et al, 2012) and annotated using Prokka 1.12 (Seeman, 2014).

Comparative Analysis

Pangenomic analysis was completed using the contigs extracted from the Genbank file which were re-annotated using Prokka 1.13.3 (Seeman, 2014). This generated GFF files that were used as input for Roary 3.12 (Page et al, 2015). Multi-locus sequence typing (MLST) was performed using online MLST databases. In particular, the *Enterococcus faecalis* MLST (<https://pubmlst.org/efaecalis/>) and *Enterococcus faecium* MLST (<https://pubmlst.org/efaecium/>) based at the University of Oxford (Jolley & Maiden, 2010) and funded by the Wellcome Trust. The phylogenetic trees were constructed based on analysis of single nucleotide variants (SNVs) of the core genes. The phylogenetic analyses were conducted using a single nucleotide variant phylogenomics (SNVPhyl) pipeline (Petkau et al, 2017) using unassembled sequence read data. The paired-end reads for Illumina sequencing of the 39 *Enterococcus spp.* genomes were aligned to the appropriate reference genome to generate read pileups (SMALT v.0.7.5; <http://www.sanger.ac.uk/science/tools/smalt-0>). The presence and absence matrices were generated using Phandango (Hadfield et al, 2018). Whole genome sequences of *E. faecalis* and *E. faecium* (Additional File 1) were also included in the analysis and were ran through the ART next-generation sequencing read simulator (Huang et al., 2012) to generate paired-end reads with length and coverage similar to the experimental dataset (2 × 300 base PE and ~50X coverage). The reads were subject to mapping quality filtering (minimum mean mapping quality score of 30) and coverage (15X minimum coverage threshold) estimations. Using a single nucleotide variant (SNV) abundance ratio of 0.75, with no SNV density filtering setting, variant calling, variant consolidation and single nucleotide variant alignment generation of the final phylogeny was run through PhyML (Guindon and Gascuel, 2003) using maximum the likelihood method. The resulting tree was visualized using interactive Tree of Life (iTOL) version 4.2.1 (<https://itol.embl.de/>). Assignment of proteins into Clusters of Orthologous Groups (COGs) was performed using the *compare genomes* function of DOE Joint Genome Institute Integrated Microbial Genomes & Microbiomes platform (Markowitz et al, 2011). Correlations were calculated using R statistical platform version 3.4.3 (R Core Team, 2013) and figures were generated using packages Hmisc (Harrell & Harrell, 2019) and corrplot (Wei et al, 2017).

Draft genome sequences of the 39 *Enterococcus spp.* were investigated for the presence of putative virulence and AMR genes, mobile gene elements, bacteriophage, and CRISPR/Cas arrays. The contigs of each draft genome were ordered based on alignment against a reference genome using progressive Mauve (Darling et al, 2010). Virulence and AMR genes were identified using Virulence Finder version 1.5 (Joensen et al, 2014) and CARD version 2.0.1 (Jia et al, 2017), respectively. Results for AMR genes were further verified using megaBLAST and hits were manually curated. Genomes were investigated for integrative conjugative elements (ICEs) by homology searches using BLAST against 466 ICEs downloaded from the ICEberg database 1.0 (Bi et al, 2012). The genomes were then analyzed for the presence of prophage using PHAST (Zhou et al, 2011). CRISPR-Cas arrays were identified using the CRISPRdb (Grissa et al, 2007). Secondary metabolites biosynthetic gene clusters were identified using the Antibiotics and Secondary Metabolite Analysis Shell (antiSMASH) version 3.0 (Weber et al, 2015).

A biomarker search was carried out with the 39 genomes from this study and an additional 59 genomes retrieved from NCBI using Neptune (Marinier et al, 2017) and a Galaxy instance from the National Microbiology Laboratory in Winnipeg, MB, Canada. The inclusion and exclusion groups are listed in Additional File 1 (Sheet 19). The cut-off score for signatures among species was 95% and the cut-off score for signatures within species from different sources was 80%. The functions related to the genes covered by each signature was identified by mapping the signatures to a reference then identifying the functions of the genes using UniProt (UniProt Consortium, 2014). The reference genomes that were used were *E. faecalis* V583 (NC_004668), *E. faecium* DO (NC_017960), and *E. casseliflavus* B72 (this study).

List Of Abbreviations

AMR: antimicrobial resistance

antiSMASH : Antibiotics and Secondary Metabolite Analysis Shell

ARG: antimicrobial resistance gene

BAF : biological aerated filter

BHI : Blood Heart Infusion

BLAST : basic local alignment search tool

CARD : Comprehensive Antimicrobial Resistance Database

CAS : conventional activated sludge

CLST : Clinical and Laboratory Standards Institute

COGs : clusters of orthologous groups

CRISPRdb : CRISPR database

CRISPR/Cas : Clustered regularly interspaced short palindromic repeats and CRISPR-associated genes

crRNAs: CRISPR RNA

EUCAST: European Committee on Antimicrobial Resistance Testing

ICE: integrated conjugative element

iTOL: interactive tree of life

MGE: mobile genetic element

PHAST: phage search tool

SNVPhyl: single nucleotide variants phylogenomics pipeline

ST: sequence type

VRE: vancomycin-resistant enterococci

WWTP: wastewater treatment plant

Declarations

Consent for Publication

Not applicable.

Ethics Approval

Not applicable.

Availability of Data and Materials

The datasets supporting the conclusions of this article are included within the article and its additional files. The genomes sequenced in this articles are available from NCBI Bioproject PRJNA524668.

Competing Interests

The authors declare that they have no competing interests.

Funding

Authors are grateful to the Natural Sciences and Engineering Research Council of Canada (NSERC) for financial support, Queen's University, Public Health Ontario and Agriculture and Agri-food Canada. HS received Ontario Graduate Student Scholarships and a Queen Elizabeth II Science and Technology

Scholarship while completing this work. Support by the Genomics Research and Development Initiative of the Government of Canada and BCRC Project FOS 10.13 are also appreciated.

Authors' Contributions

This work was part of a study that contributed to a PhD thesis by HS. HS, ROP, and RZ designed the analysis. HS extracted the DNA and prepared samples for sequencing. NG sequenced the isolates. KKA contributed funding for the sequencing of the isolates. RZ performed the assembly, annotation, and phylogenetic analyses of genomes. HS and ROP performed the bioinformatics analysis. HS analyzed the results and drafted the manuscript. TAM RSB, AM and SNL provided funding or equipment necessary to complete the work. AM and RSB supervised HS. All authors revised the manuscript and approved the final manuscript.

Acknowledgements

Thank you to Dr. Chad Liang for revising this manuscript

References

Arias, C. A., & Murray, B. E. Emergence and management of drug-resistant enterococcal infections. *Expert Review of Anti-infective Therapy*. 2008; 6, 637-655.

Ashbolt, N. J., Amézquita, A., Backhaus, T., Borriello, P., Brandt, K. K., Collignon, P., Coors, A., Finley, R., Gaze, W. H., Heberer, T., Lawrence, J. R., Larsson, D. G., McEwen, S. A., Ryan, J. J., Schonfeld, J., Silley, P., Snape, J. R., van den Eede, C., & Topp, E. Human health risk assessment (HHRA) for environmental development and transfer of antibiotic resistance. *Environmental Health Perspectives*. 2013; 121, 993.

Baker-Austin, C., Wright, M. S., Stepanauskas, R., & McArthur, J. V. Co-selection of antibiotic and metal resistance. *Trends in microbiology*. 2006; 14(4), 176-182.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*. 2012; 19, 455-477.

Barrangou, R., & Horvath, P. The CRISPR system protects microbes against phages, plasmids. *Microbe*. 2009; 4, 224-230.

Bentley, S. D., Chater, K. F., Cerdeno-Tarraga, A. M., Challis, G. L., Thomson, N. R., James, K. D., Harris, D. E., Quail, M. A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C. W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C. H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabinowitsch, E., Rajandream, M. A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B. G., Parkhill

- J., & Hopwood, D. A. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3 (2). *Nature*. 2002; *417*, 141.
- Beukers, A. G., Zaheer, R., Goji, N., Amoako, K. K., Chaves, A. V., Ward, M. P., & McAllister, T. A. Comparative genomics of *Enterococcus spp.* isolated from bovine feces. *BMC Microbiology*. 2017; *17*, 52.
- Bi, D., Xu, Z., Harrison, E. M., Tai, C., Wei, Y., He, X., Jia, S., Deng, Z., Rajakumar, K., & Ou, H. Y. ICEberg: a Web-based resource for integrative and conjugative elements found in Bacteria. *Nucleic Acids Resesearch*. 2012; *40*, D621–D626.
- Blanch, A. R., Caplin, J. L., Iversen, A., Kühn, I., Manero, A., Taylor, H. D., & Vilanova, X. Comparison of enterococcal populations related to urban and hospital wastewater in various climatic and geographic European regions. *Journal of Applied Microbiology*. 2003; *94*, 994-1002.
- Boehm, A. B., & Salubrious, L. M. Enterococci as indicators of environmental fecal contamination. Gilmore MS, Clewell DB, Ike Y, et al., editors. *Enterococci: From Commensals to Leading Causes of Drug Resistant Infection* [Internet]. Boston: Massachusetts Eye and Ear Infirmary; 2014. Retrieved from <
<https://www.ncbi.nlm.nih.gov/books/NBK190421/>>
- Bondy-Denomy, J., Pawluk, A., Maxwell, K. L., & Davidson, A. R. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature*. 2013; *493*, 429.
- Booth, M. C., Bogie, C. P., Sahl, H. G., Siezen, R. J., Hatter, K. L., & Gilmore, M. S. Structural analysis and proteolytic activation of *Enterococcus faecalis* cytolysin, a novel lantibiotic. *Molecular Microbiology*. 1996; *21*, 1175-1184.
- Caplin, J. L., Hanlon, G. W., & Taylor, H. D. Presence of vancomycin and ampicillin-resistant *Enterococcus faecium* of epidemic clonal complex-17 in wastewaters from the south coast of England. *Environmental Microbiology*. 2008; *10*, 885-892.
- Chang, Y. M., Chen, C. M., Ko, T. P., Chang-Chien, M. W., & Wang, A. J. Structural analysis of the antibiotic-recognition mechanism of MarR proteins. *Acta Crystallographica Section D: Biological Crystallography*. 2013; *69*, 1138-1149.
- Chen, L., Ruan, Z., Wu, Q., Zhou, T., & Li, C. Relationship between ermB gene and transposons Tn1545 and Tn917 in *Enterococcus faecalis*. *Chinese Journal of Nosocomiology*. 2010; *20*(6), 763-766.
- Darling, A. E., Mau, B., & Perna, N. T. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PloS one*. 2010; *5*, e11147.
- Duerkop, B. A., Palmer, K. L., & Horsburgh, M. J. Enterococcal bacteriophages and genome defense. 2014. In: Gilmore, M. S., Clewell, D.B., Ike, Y., Shankar, N., Ed. *Enterococci: from commensals to leading cause of drug resistant infection*. Boston: Massachusetts Eye and Ear Infirmary. 2014; [online]<
<https://www.ncbi.nlm.nih.gov/books/NBK190419/>>

Faron, M. L., Ledebøer, N. A., & Buchan, B. W. Resistance mechanisms, epidemiology, and approaches to screening for vancomycin-resistant *Enterococcus* in the health care setting. *Journal of Clinical Microbiology*. 2016; *54*, 2436-2447.

Grissa, I., Vergnaud, G., & Pourcel, C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Research*. 2007; *35*, W52-W57.

Grove, A. MarR family transcription factors. *Current Biology*. 2013; *23*, R142-R143.

Guindon, S., & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*. 2003; *52*, 696-704.

Hadfield, J., Croucher, N.J., Goater, R. J., Abudahab, K., Aanensen, D. M., Harris, S. R.. Phandango: an interactive viewer for bacterial population genomics, *Bioinformatics*, Volume 34, Issue 2, 2018; 292–293, <https://doi.org/10.1093/bioinformatics/btx610>

Hammerum, A. M. Enterococci of animal origin and their significance for public health. *Clinical Microbiology and Infection*. 2012; *18*, 619-625.

Harrell Jr, F. E., & Harrell Jr, M. F. E. Package 'Hmisc'. *CRAN2018*, 2019; 235-6.

Hegemann, J. D., Zimmermann, M., Xie, X., & Marahiel, M. A. Lasso peptides: an intriguing class of bacterial natural products. *Accounts of Chemical Research*. 2015; *48*, 1909-1919.

Hölzel, C. S., Harms, K. S., Küchenhoff, H., Kunz, A., Müller, C., Meyer, K., Schwaiger, K., & Bauer, J. Phenotypic and genotypic bacterial antimicrobial resistance in liquid pig manure is variously associated with contents of tetracyclines and sulfonamides. *Journal of Applied Microbiology*. 2010; *108*, 1642-1656.

Huang, W., Li, L., Myers, J.R., Marth, G.T. ART: a next-generation sequencing read simulator. *Bioinformatics*. 2012; *28*, 593-594.

Jia, B., Raphenya, A.R., Alcock, B., Waglechner, N., Guo, P., Tsang, K. K., Lago, B. A., Dave, B. M., Pereira, S., Sharma, A. N., Doshi, S., Courtot, M., Lo, R., Williams, L. E., Frye, J. G., Elsayegh, T., Sardar, D., Westman, E. L., Pawlowski, A. C., Johnson, T. A., Brinkman, F. S., Wright, G. D., & McArthur, A. G. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Research*. 2017; *45*, D566–D573. doi: 10.1093/nar/gkw1004

Joensen, K. G., Scheutz, F., Lund, O., Hasman, H., Kaas, R. S., Nielsen, E. M., & Aarestrup, F. M. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *Journal of Clinical Microbiology*. 2014; *52*, 1501-1510.

Jolley, K. A., & Maiden, M. C. BIGSdb: scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics*. 2010; *11*, 595.

- Kell, D. B., Kaprelyants, A. S., & Grafen, A. Pheromones, social behaviour and the functions of secondary metabolism in bacteria. *Proc. Natl Acad Sci USA*. 1995; *71*, 2843-2847.
- Klein, G. Taxonomy, ecology and antibiotic resistance of enterococci from food and the gastro-intestinal tract. *International Journal of Food Microbiology*. 2003; *88*, 123-131.
- Knerr, P. J., & Van Der Donk, W. A. Discovery, biosynthesis, and engineering of lantipeptides. *Annual Review of Biochemistry*. 2012; *81*, 479-505.
- Konstantinidis, K. T., & Tiedje, J. M. Trends between gene content and genome size in prokaryotic species with larger genomes. *Proceedings of the National Academy of Sciences*. 2004; *101*, 3160-3165.
- Loong, S. K., Che Mat Seri, N. A. A., Mahfodz, N. H., Ahmad Nasrah, S. N., Akbar, S. Z., & AbuBakar, S. Research Note A Report of Vancomycin-susceptible, Teicoplanin-resistant *Enterococcus faecalis* ST6 in Malaysia. *Tropical Biomedicine*. 2016; *33*, 577-582.
- Maraccini, P. A., Ferguson, D. M., & Boehm, A. B. Diurnal variation in Enterococcus species composition in polluted ocean water and a potential role for the enterococcal carotenoid in protection against photoinactivation. *Appl. Environ. Microbiol.* 2012; *78*(2), 305-310.
- Marinier E, Zaheer R, Berry C, Weedmark KA, Domaratzki M, Mabon P, Knox NC, Reimer AR, Graham MR, Chui L, Patterson-Fortin L. Neptune: a bioinformatics tool for rapid discovery of genomic variation in bacterial populations. *Nucleic acids research*. 2017 Aug 17;45(18):e159-.
- Markowitz, V. M., Chen, I. M. A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., Ratner, A., Jacob, B., Huang, J., Williams, P., Huntemann, M., Anderson, I., Mavromatis, K., Ivanova, N. N., & Kyrpides, N. C. IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Research*. 2011; *40*, D115-D122.
- Martins Da Costa, P., Vaz-Pires, P., & Bernardo, F. Antimicrobial resistance in Enterococcus spp. isolated in inflow, effluent and sludge from municipal sewage water treatment plants. *Water Research*. 2006a; *40*, 1735-1740.
- Martins da Costa, P. M., Vaz-Pires, P. M., & Bernardo, F. M. Antibiotic resistance of Enterococcus spp. isolated from wastewater and sludge of poultry slaughterhouses. *Journal of Environmental Science and Health Part B*. 2006b; *41*, 1393-1403.
- Mazaheri Nezhad Fard, R. , Barton, M. D., & Heuzenroeder, M. W. Novel bacteriophages in enterococcus spp. *Current Microbiology*. 2010; *60*, 400-406.
- Miller, W. R., Munita, J. M., & Arias, C. A. Mechanisms of antibiotic resistance in enterococci. *Expert Review of Anti-infective Therapy*. 2014; *12*, 1221-1236.

- Müller, T., Ulrich, A., Ott, E. M., & Müller, M. Identification of plant-associated enterococci. *Journal of Applied Microbiology*. 2001; *91*, 268-278.
- Munck, C., Albertsen, M., Telke, A., Ellabaan, M., Nielsen, P. H., & Sommer, M. O. Limited dissemination of the wastewater treatment plant core resistome. *Nature Communications*. 2015; *6*, 8452.
- Paduch, R., Kandefer-Szerszeń, M., Trytek, M., & Fiedurek, J. Terpenes: substances useful in human healthcare. *Archivum Immunologiae et Therapiae Experimentalis*. 2007; *55*, 315.
- Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T., ... & Parkhill, J. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*. 2015; *31*(22), 3691-3693.
- Palmer, K. L., & Gilmore, M. S. 2010. Multidrug-resistant enterococci lack CRISPR-cas. *MBio*. 2010; *1*, e00227-10.
- Palmer, K. L., Godfrey, P., Griggs, A., Kos, V. N., Zucker, J., Desjardins, C., Cerqueira, G., Gevers, D., Walker, S., Wortman, J., Feldgarden, M., Haas, B., Birren, B., & Gilmore, M. S. Comparative genomics of enterococci: variation in *Enterococcus faecalis*, clade structure in *E. faecium*, and defining characteristics of *E. gallinarum* and *E. casseliflavus*. *MBio*. 2012; *3*, e00318-11.
- Petkau, A., Stuart-Edwards, M., Stothard, P., & Van Domselaar, G. Interactive microbial genome visualization with GView. *Bioinformatics*. 2010; *26*, 3125-3126.
- Petkau, A., Mabon, P., Sieffert, C., Knox, N. C., Cabral, J., Iskander, M., Iskander, M., Weedmark, K., Zaheer, R., Katz, L. S., Nadon, C., Reimer, A., Taboada, E., Beiko, R. G., Hsiao, W., Brinkman, F., Graaham, M., & van Domselaar, G. SNVPhyl: a single nucleotide variant phylogenomics pipeline for microbial genomic epidemiology. *Microbial Genomics*. 2017; *3*.
- Qin, X., Galloway-Peña, J. R., Sillanpää, J., Roh, J. H., Nallapareddy, S. R., Chowdhury, S., Bourgoigne, A., Choudhury, T., Muzny, D. M., Buhay, C. J., Ding, Y., Dugan-Rocha, S., Liu, W., Kovar, C., Sodergren, E., Highlander, S., Petrosino, J. F., Worley, K. C., Gibbs, R. A., Weinstock, G. M., & Murray, B. E. Complete genome sequence of *Enterococcus faecium* strain TX16 and comparative genomic analysis of *Enterococcus faecium* genomes. *BMC Microbiology*. 2012; *12*, 135.
- R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Rendueles, O., Garcia-Garcerà, M., Néron, B., Touchon, M., & Rocha, E. P. Abundance and co-occurrence of extracellular capsules increase environmental breadth: Implications for the emergence of pathogens. *PLoS pathogens*. 2017; *13*, e1006525.
- Rice, L. B. Tn916 family conjugative transposons and dissemination of antimicrobial resistance determinants. *Antimicrobial Agents and Chemotherapy*. 1998; *42*, 1871-1877.

- Rice, L. B., & Carias, L. L. Transfer of Tn5385, a composite, multiresistance chromosomal element from *Enterococcus faecalis*. *Journal of Bacteriology*. 1998; *180*, 714-721.
- Ricker, N., Qian, H., & Fulthorpe, R. R. The limitations of draft assemblies for understanding prokaryotic adaptation and evolution. *Genomics*. 2012; *100*, 167-175.
- Roberts, M. C. Resistance to tetracycline, macrolide-lincosamide-streptogramin, trimethoprim, and sulfonamide drug classes. *Molecular Biotechnology*. 2002; *20*, 261-283.
- Roberts, A. P., & Mullany, P. Tn 916-like genetic elements: a diverse group of modular mobile elements conferring antibiotic resistance. *FEMS Microbiology Reviews*. 2011; *35*, 856-871.
- Roca, I., Akova, M., Baquero, F., Carlet, J., Cavaleri, M., Coenen, S., Cohen, J., Findlay, D., Gyssens, I., Heure, O. E., Kahlmeter, G., Kruse, H., Laxminarayan, R., Liebana, E., Lopez-Cerero, L., MacGowen, A., Martins, M., Rodriguez-Bano, J., Rolain, J. M., Segavia, C., Sigauque, B., Tacconelli, E., Wellington, E. & Vila, J. The global threat of antimicrobial resistance: science for intervention. *New Microbes and New Infections*. 2015; *6*, 22-29.
- Roszak, D. B., & Colwell, R. R. Survival strategies of bacteria in the natural environment. *Microbiological Reviews*. 1987; *51*, 365.
- Ruiz-Garbajosa, P., Bonten, M. J., Robinson, D. A., Top, J., Nallapareddy, S. R., Torres, C., Coque, T. M., Canton, R., Baquero, F., Murray, B. E., del Campo, R. & Willems, R. J. L. Multilocus sequence typing scheme for *Enterococcus faecalis* reveals hospital-adapted genetic complexes in a background of high rates of recombination. *Journal of Clinical Microbiology*. 2006; *44*, 2220-2228.
- Sanderson, H., Fricker, C., Brown, R. S., Majury, A., & Liss, S. N. Antibiotic resistance genes as an emerging environmental contaminant. *Environmental Reviews*. 2016; *24*, 205-218.
- Sanderson, H., Ortega-Polo, R., McDermott, K., Zaheer, R., Brown, R. S., Majury, A., McAllister, T. A. & Liss, S. N. Comparison of biochemical and genotypic speciation methods for vancomycin-resistant enterococci isolated from urban wastewater treatment plants. *Journal of microbiological methods*. 2019; *161*, 102-110.
- Schöner, T. A., Gassel, S., Osawa, A., Tobias, N. J., Okuno, Y., Sakakibara, Y., Shindo, K., Sandmann, G., & Bode, H. B. Aryl polyenes, a highly abundant class of bacterial natural products, are functionally related to antioxidative carotenoids. *ChemBioChem*. 2016; *17*, 247-253.
- Shmakov, S. A., Sitnik, V., Makarova, K. S., Wolf, Y. I., Severinov, K. V., & Koonin, E. V. The CRISPR spacer space is dominated by sequences from species-specific mobilomes. *MBio*. 2017; *8*(5), e01397-17.
- Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014; *30*, 2068-2069.

Shaw, J. H., & Clewell, D. B. Complete nucleotide sequence of macrolide-lincosamide-streptogramin B-resistance transposon Tn917 in *Streptococcus faecalis*. *Journal of Bacteriology*. 1985; *164*, 782-796.

Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warrener, P., Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J., Lagrou, M., Garber, R. L., Goltry, L., Tolentino, E., Westbrook-Wadman, S., Yuan, Y., Brody, L. L., Coulter, S. N., Folger, K. R., Kas, A., Larbig, K., Lim, R., Smith, K., Spencer, D., Wong, G. K. S., Wu, Z., Paulsen, I. T., Reizer, J., Saier, M. H., Hancock, R. E. W., Lory, S., & Olson, M. V. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature*. 2000; *406*: 959–964. doi:10.1038/35023079. PMID:10984043.

Strateva, T., Atanasova, D., Savov, E., Petrova, G., & Mitov, I. Incidence of virulence determinants in clinical *Enterococcus faecalis* and *Enterococcus faecium* isolates collected in Bulgaria. *Brazilian Journal of Infectious Diseases*. 2016; *20*, 127-133.

Tao, P., Wu, X., & Rao, V. Unexpected evolutionary benefit to phages imparted by bacterial CRISPR-Cas9. *Science Advances*. 2018; *4*, eaar4134.

Tatusov, R. L., Natale, D. A., Garkavtsev, I. V., Tatusova, T. A., Shankavaram, U. T., Rao, B. S., Kiryutin, B., Galperin, M. Y., Fedorova, N. D., & Koonin, E. V. The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Research*. 2001; *29*, 22-28.

UniProt Consortium. UniProt: a hub for protein information. *Nucleic acids research*. 2014 Oct 27; *43*(D1):D204-12.

van Elsas, J., & Bailey, M. J. The ecology of transfer of mobile genetic elements. *FEMS Microbiology Ecology*. 2002; *42*: 187–197. doi:10.1111/j.1574-6941.2002.tb01008.x. PMID:19709278.

van Schaik, W., Top, J., Riley, D. R., Boekhorst, J., Vrijenhoek, J. E., Schapendonk, C. M., Hendrickx, A. P. A., Nijman, I. J., Bonten, M. J. M., Tettelin, H., & Willems, R. J. L. Pyrosequencing-based comparative genome analysis of the nosocomial pathogen *Enterococcus faecium* and identification of a large transferable pathogenicity island. *BMC Genomics*. 2010; *11*, 239.

van Tyne, D., & Gilmore, M. S. Friend turned foe: evolution of enterococcal virulence and antibiotic resistance. *Annual Review of Microbiology*. 2014; *68*, 337-356.

Varela, A. R., Ferro, G., Vredenburg, J., Yanik, M., Vieira, L., Rizzo, L., Lameiras, C., & Manaia, C. M. Vancomycin resistant enterococci: from the hospital effluent to the urban wastewater treatment plant. *Science of the Total Environment*. 2013; *450*, 155-161.

Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H. U., Bruccoleri, R., Lee, S. Y., Fischbach, M. A., Muller, R., Wohlleben, W., Breitling, R., Takano, E., & Medema, M. H. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Research*. 2015; *43*, W237-W243.

Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y., & Zemla, J. (2017). Package 'corrplot'. *Statistician*. 2017; 56, 316-324.

Weigand, M. R., Ashbolt, N. J., Konstantinidis, K. T., & Santo Domingo, J. W. Genome sequencing reveals the environmental origin of enterococci and potential biomarkers for water quality monitoring. *Environmental science & technology*. 2014; 48(7), 3707-3714.

Yang, S. C., Lin, C. H., Sung, C. T., & Fang, J. Y. Antibacterial activities of bacteriocins: application in foods and pharmaceuticals. *Frontiers in Microbiology*. 2014; 5, 241.

Zaheer, R., Yanke, L. J., Church, D., Topp, E., Read, R. R., & McAllister, T. A. High-throughput species identification of enterococci using pyrosequencing. *Journal of Microbiological Methods*. 2012; 89, 174-178.

Zhong, Z., Zhang, W., Song, Y., Liu, W., Xu, H., Xi, X., Menghe, B., Zhang, H., & Sun, Z. Comparative genomic analysis of the genus *Enterococcus*. *Microbiological Research*. 2017; 196, 95-105.

Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., & Wishart, D. S. PHAST: a fast phage search tool. *Nucleic Acids Research*. 2011; 39, W347-W352.

Additional File Legend

Additional File 1

Format: Excel

Title of Data: Supplementary Details on Genomic Comparisons

Description: Sheet 1: Complete List of Genomes Used; Sheet 2: Genome Characteristics of Sequenced Genomes; Sheet 3: COG Analysis; Sheet 4: COG Pivot Chart; Sheet 5: COG graphs; Sheet 6: COG correlation and p values; Sheet 7: AMR phenotypic profiles and breakpoints; Sheet 8: AMR genes detected; Sheet 9: Virulence Factors Raw Output; Sheet 10: Virulence Factors by Species; Sheet 11: Virulence Factors Summary; Sheet 12: Phage Detected; Sheet 13: CRISPRfinder Output; Sheet 14: CRISPRfinder Confirmed Systems; Sheet 15: CRISPRdb BLAST Results; Sheet 16: CRISPR Summary Table; Sheet 17: ICEberg Raw Results; Sheet 18: antiSMASH results; Sheet 19: Neptune Biomarker Combinations

Figures

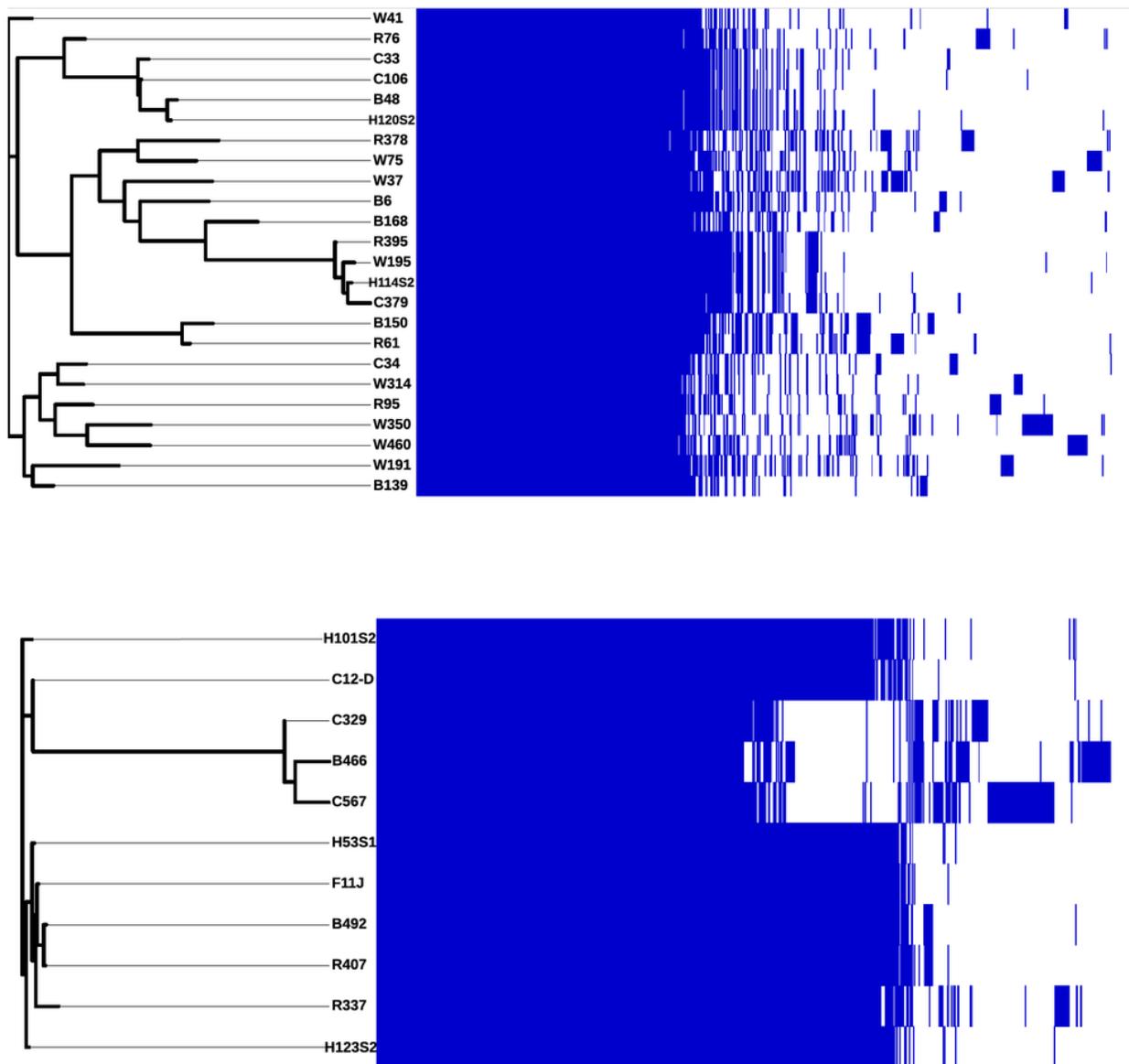


Figure 1

Enterococcus faecalis (a) and *Enterococcus faecium* (b) pan-genome illustrated as a matrix with the core SNP tree of the strains on the left and a presence (blue) and absence (white) matrix of core and accessory genes.

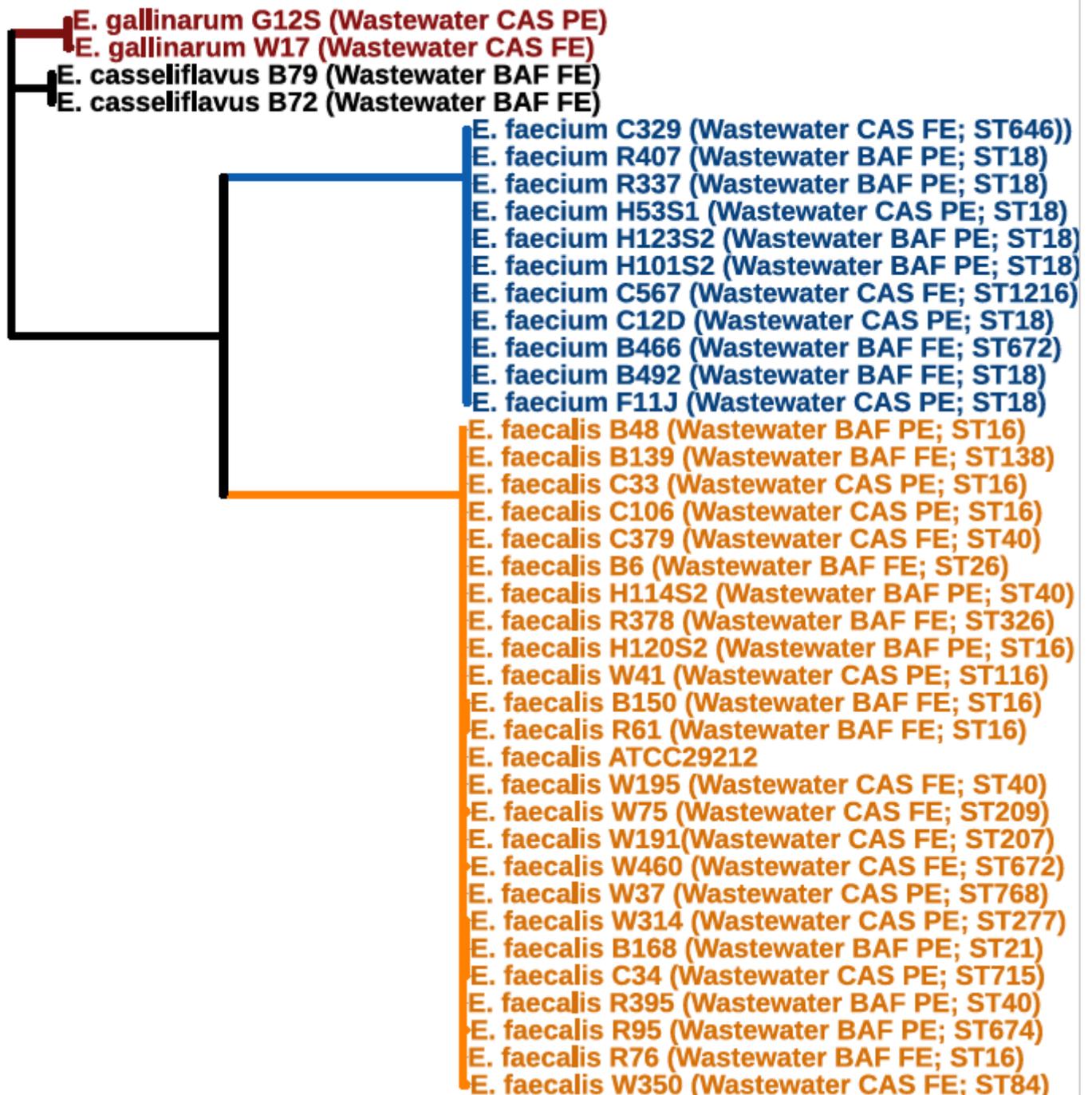


Figure 2

Phylogenetic Tree of all Enterococcus spp. isolated from wastewater using Enterococcus faecalis ATCC 29212 as the reference genome.

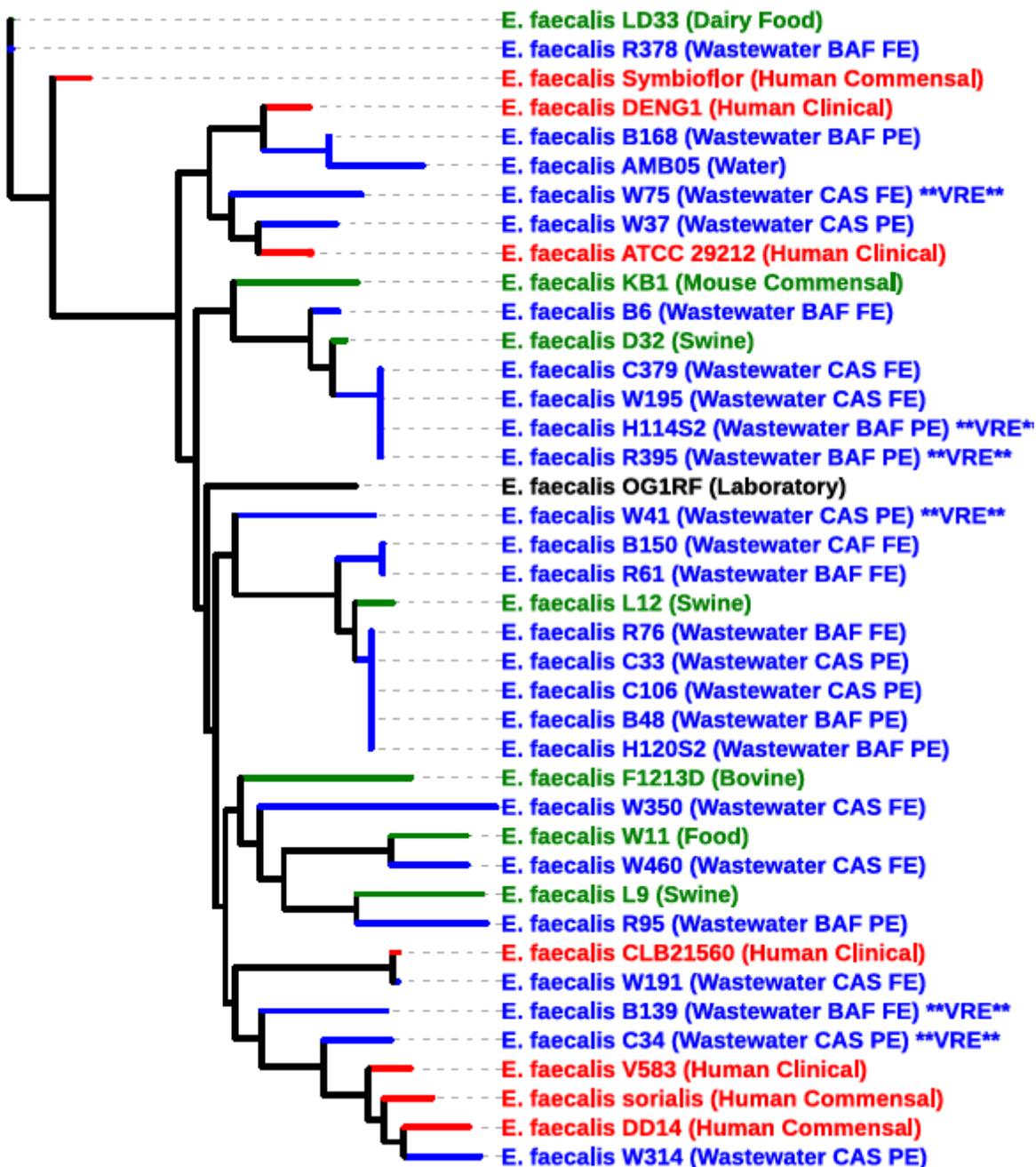


Figure 3

Phylogenetic tree of *Enterococcus faecalis* genome sequences from the present study and complete genome sequences from the NCBI GenBank database based on analysis of single-nucleotide variants (SNVs) of the core genes. *Enterococcus faecalis* ATCC29212 was used as the reference genome. Origin of Isolates are as indicated in the figures and are grouped by colour into clinical (red), agricultural/food (green) and wastewater/water (blue) groups.

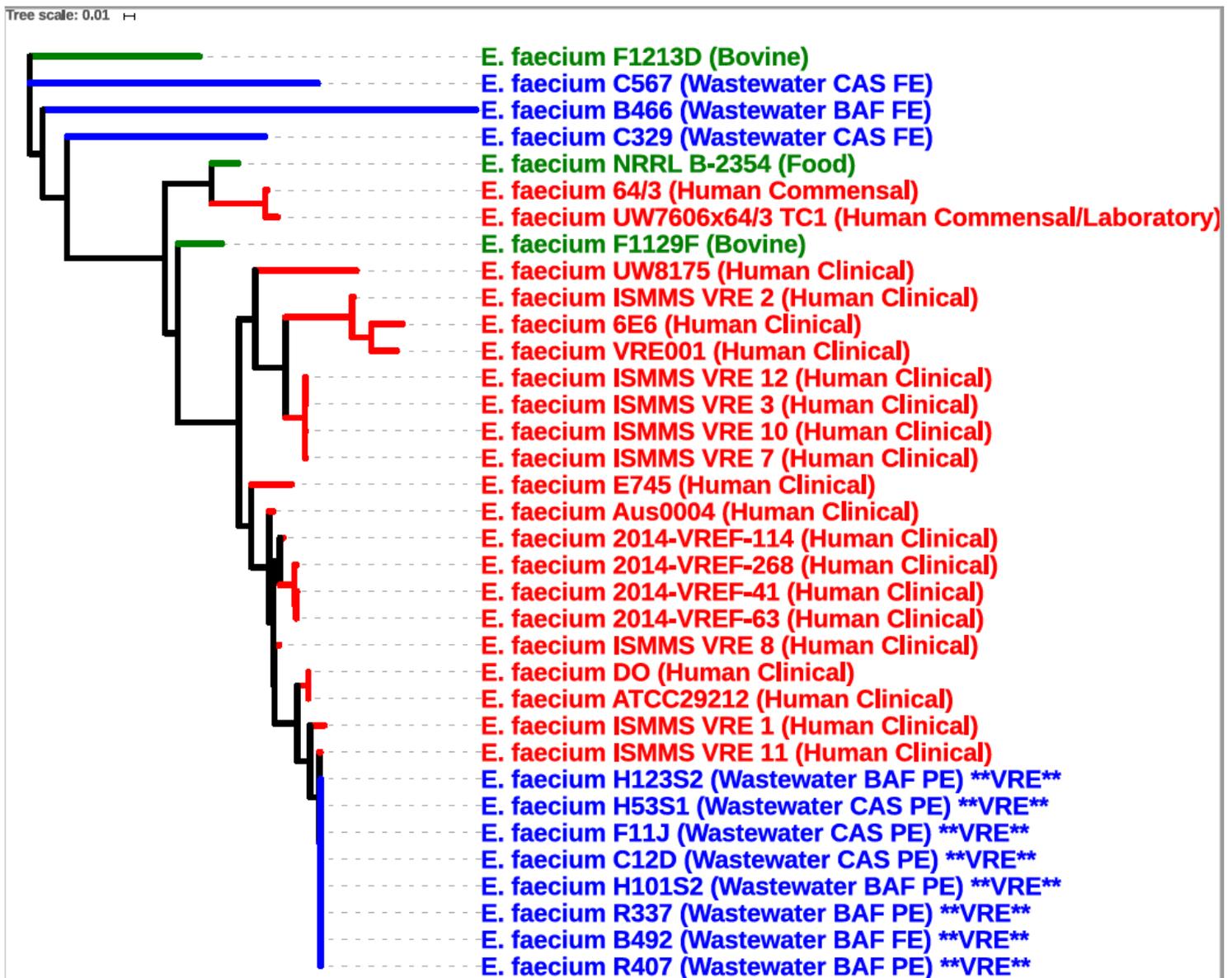


Figure 4

Phylogenetic tree of *Enterococcus faecium* genome sequences from the present study and genome sequences from the NCBI GenBank database based on analysis of single-nucleotide variants (SNVs) of the core genes. *Enterococcus faecium* DO served as the reference genome. Origin of isolates are as indicated in the figures and are grouped by colour into clinical (red), agricultural/food (green) and wastewater/water (blue) groups

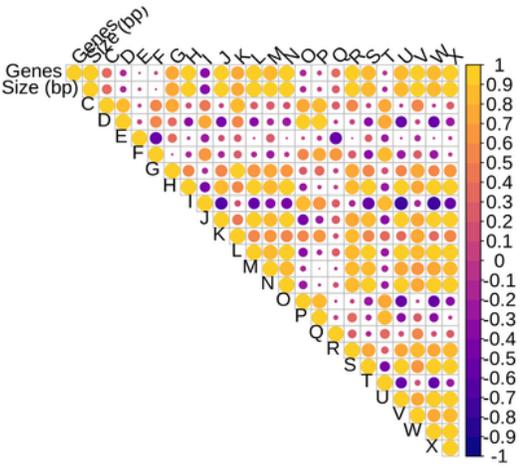
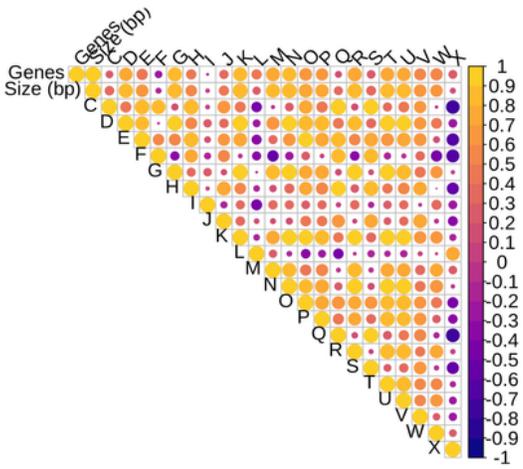
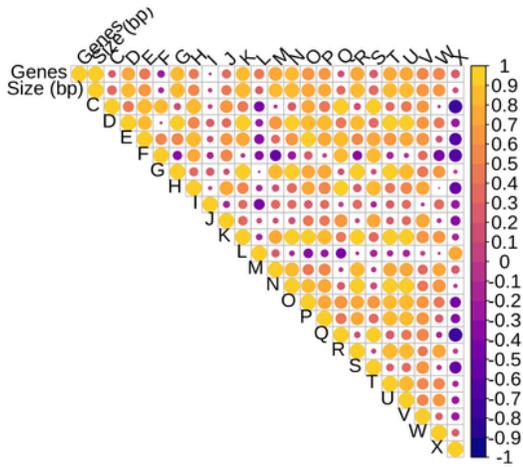


Figure 5

Correlatin plots showing the correlations between different COG categories, genome size and number of genes in all of the pooled *Enterococcus* species (a), *E. faecalis*, (b) and *E. faecium* (c).

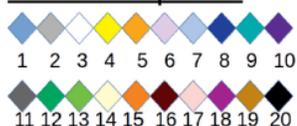
Functional CRISPR Arrays

E. faecalis B6	◆ ₁₀
E. faecalis B48	◆ ₉
E. faecalis B139*	◆ ₁₇
E. faecalis C106	◆ ₈
E. faecalis C33	◆ ₉
E. faecalis C379	◆ ₁₂
E. faecalis H114S2	◆ ₁₂
E. faecalis H120S2	◆ ₉
E. faecalis R76	◆ ₈
E. faecalis R95	◆ ₁₉
E. faecalis R395	◆ ₁₂
E. faecalis W195	◆ ₁₂
E. faecalis W460	◆ ₁₁

Orphan CRISPR Arrays

E. casseliflavus B72	◆ ₅	E. faecalis W37	◆ ₄
E. casseliflavus B79	◆ ₅	E. faecalis W41*	◆ ₁₄ ◆ ₆
E. faecalis B48	◆ ₉ ◆ ₁₀	E. faecalis W75	◆ ₅
E. faecalis B150	◆ ₄	E. faecalis W195	◆ ₄
E. faecalis B168*	◆ ₄	E. faecalis W350	◆ ₈
E. faecalis C106	◆ ₁₀	E. faecalis W460	◆ ₁₁ ◆ ₉ ◆ ₄
E. faecalis C33	◆ ₁₀	E. faecium B492	◆ ₃
E. faecalis C379	◆ ₄	E. faecium C12D	◆ ₃
E. faecalis H114S2	◆ ₄	E. faecium C567	◆ ₃
E. faecalis H120S2	◆ ₁₀	E. faecium F11J	◆ ₃
E. faecalis R61	◆ ₄	E. faecium H101S2	◆ ₃
E. faecalis R76	◆ ₉	E. faecium H53S1	◆ ₃
E. faecalis R95*	◆ ₇	E. faecium R337	◆ ₃
E. faecalis R378	◆ ₉	E. faecium R407	◆ ₃
E. faecalis R395	◆ ₄		

Direct Repeats



No CRISPR Arrays

E. faecalis: C34, W191, W314
 E. faecium: C329*, B466, H123S2
 E. gallinarum: G12S, W17

Figure 6

Pictorial of CRISPR-Cas arrays identified in the analysis of wastewater Enterococcus spp (n = 39) genomes. (A) Functional CRISPR arrays with the organization of direct repeats (diamonds) and spacers (numbers). (B) Orphan CRISPR arrays direct repeat and spacer organization. (C) Legend for numbered direct repeats and the genomes with no CRISPR arrays. The outlined genomes are those that contained both functional CRISPR arrays and prophage. The (*) represents genomes that contained no prophage.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [WGScomparativegenomicssupplementarymaterialOct152019.xlsx](#)