LEBANESE AMERICAN UNIVERSITY

Genome profiling of vancomycin-resistant enterococcal clinical Isolates from Lebanon

By

Balig V. Panossian

A thesis

Submitted in partial fulfillment of the requirements for

the degree of Master of Science in Molecular Biology

School of Arts and Sciences

May 2019

LAU
الجامعة اللبنانية الأميركية
Lebanese American University

THESIS APPROVAL FORM

Student Name: Balig Panossian I.D. #: 201304460

Thesis Title: Genome Profiling of vancomycin-resistant enterococcal clinical isolates from lebanon

Program: Molecular Biology

Department: Natural Sciences

School: Arts and Sciences

The undersigned certify that they have examined the final electronic copy of this thesis and approved it in Partial Fulfillment of the requirements for the degree of:

Master of Science in the major of Molecular Biology

Thesis Advisor's Name: Dr Sima Tokajian, Dr Georges Khazen

Signature [redacted] Date: 9 / 5 / 19 (Day / Month / Year)

Committee Member's Name: SANDRA RIZK

Signature: [redacted] Date: 9 / 5 / 19 (Day / Month / Year)

Committee Member's Name: CHRISTIAN KHALIL

Signature: [redacted] Date: 9 / 5 / 19 (Day / Month / Year)

LAU
الجامعة اللبنانية الأميركية
Lebanese American University

THESIS COPYRIGHT RELEASE FORM

LEBANESE AMERICAN UNIVERSITY NON-EXCLUSIVE DISTRIBUTION LICENSE

By signing and submitting this license, you (the author(s) or copyright owner) grants the Lebanese American University (LAU) the non-exclusive right to reproduce, translate (as defined below), and/or distribute your submission (including the abstract) worldwide in print and electronic formats and in any medium, including but not limited to audio or video. You agree that LAU may, without changing the content, translate the submission to any medium or format for the purpose of preservation. You also agree that LAU may keep more than one copy of this submission for purposes of security, backup and preservation. You represent that the submission is your original work, and that you have the right to grant the rights contained in this license. You also represent that your submission does not, to the best of your knowledge, infringe upon anyone's copyright. If the submission contains material for which you do not hold copyright, you represent that you have obtained the unrestricted permission of the copyright owner to grant LAU the rights required by this license, and that such third-party owned material is clearly identified and acknowledged within the text or content of the submission. IF THE SUBMISSION IS BASED UPON WORK THAT HAS BEEN SPONSORED OR SUPPORTED BY AN AGENCY OR ORGANIZATION OTHER THAN LAU, YOU REPRESENT THAT YOU HAVE FULFILLED ANY RIGHT OF REVIEW OR OTHER OBLIGATIONS REQUIRED BY SUCH CONTRACT OR AGREEMENT. LAU will clearly identify your name(s) as the author(s) or owner(s) of the submission, and will not make any alteration, other than as allowed by this license, to your submission.

Name: Balig Viken Panossian

Signature: [redacted]

Date: 09 / 05 / 2019
Day Month Year

LAU
الجامعة اللبنانية الأميركية
Lebanese American University

PLAGIARISM POLICY COMPLIANCE STATEMENT

I certify that:

1. I have read and understood LAU's Plagiarism Policy.
2. I understand that failure to comply with this Policy can lead to academic and disciplinary actions against me.
3. This work is substantially my own, and to the extent that any part of this work is not my own I have indicated that by acknowledging its sources.

Name: Balig Viken Panossian

Signature: [redacted]

Date: 09 / 05 / 2019
Day Month Year

ACKNOWLEDGMENTS

I would like to express my gratitude to my mentor Dr. Sima Tokajian for guiding me and encouraging me to pursue research in the field of microbial genomics. I will forever be grateful for her patience and devotion that improved me both as a scientist and as an individual in the past three years.

I am very proud to call myself a student of Dr Georges Khazen who opened up the world of bioinformatics and guided me as I took my baby steps into this amazing field. I was, and still am, awestruck by his calm and composed approach to solving even the most difficult problems.

To the microbial genomics family especially Tamara, Elie, Harout, Leila, Sahar, Jennifer, Georgi, Yara, Dana, Ghiwa, both gram positive and negative Saras, Maria, Asalia, and Rita. Thank you for all the smiles, all the jokes, all the help and support you gave to make these few years unforgettable.

A special thank you to Dr George Sadaka, the first person to encourage my curiosity and reinforce my creative inquiries starting from my first semester at LAU. Science put the world of knowledge in me, but you put me in the world of knowledge.

Finally, I would like to show my gratitude to my parents who supported me and raised me up to become the person I am today, as well as my friends who made sure I had some fun alongside all my studies. Thank you Marilyne, Christ, Missak, and Serouj for keeping me sane during times of stress and hard work.

I dedicate this work to my beloved sister, Vicky Panossian.

Genome profiling of vancomycin-resistant enterococcal clinical Isolates from Lebanon

Balig V. Panossian

Abstract

Vancomycin resistant enterococci (VRE) are a considerable burden on treatment, and infection and transmission control. Studying the intricacies on the whole-genome single nucleotide level is more informative than routine phenotypic analysis. In this study, 21 clinical isolates of enterococci were first identified by MALDI-TOF, and disk diffusion and E-test were used to determine antimicrobial susceptibility profiles. Whole-genome sequencing (WGS) was conducted using Illumina MiSeq platform. Resistance genes, ST-types, wgSNPs, cgSNPs, pan-genomes of coding and non-coding regions, as well as recombination hotspots were evaluated for all the isolates using different bioinformatic approaches. The VRE isolates clustered under two main ST types (ST80, ST203) with some isolates having very similar whole-genome average nucleotide identities (wgANI). Pan genome analysis revealed that the isolates were much more unique in their protein coding sequences and showed pronounced heterogeneity as a population. A closer look at the resistance determinants, in particular, the vancomycin resistance genes flanked by transposon fragments and recombination loci, revealed uniquely evolved

features that emphasized the polyclonal spread of these pathogens. The vancomycin-susceptible isolates also showed unique genomic features some of which had closely related genomic profiles as that of the vancomycin-resistant clones of similar ST types. A notable case was EF6 (vancomycin-susceptible), which was almost identical (99.91% average nucleotide identity) to EF15 (vancomycin resistant) except for the vancomycin resistance cassette. A few isolates were too diverse for batched intergenomic comparative analyses, and one isolate harbored additional antibiotic resistance determinants uncommon in the other enterococci.

This study provided the first insights on the genomic diversity of vancomycin-resistant enterococci circulating in Lebanon. Our results demonstrated how an end-to-end whole-genome sequence-based outbreak surveillance, and resistance profiling scheme could help in differentiating between closely related VRE isolates, whereas phenotypic or even classical molecular approaches fail to highlight these complexities.

Key Words: *Enterococcus faecalis, Enterococcus faecium,* Whole genome sequencing, Antibiotic Resistance, Vancomycin, Single Nucleotide Polymorphisms (SNP).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

pH	potential of Hydrogen
DNA	deoxyribonucleic acid
rRNA	ribosomal rybonucleic acid
RAPD	randomly amplified polymorphic DNA
sodA	manganese-dependent superoxide dismutase
efaA	enterococcus protein A
cpn60	chaperonin 60
ddl	D-ala:D-ala ligase
CC17	clonal complex 17
VRE	Vancomycin resistant enterococci
Tn	transposon
WGS	whole genome sequencing
ST	sequence type
BLAST	Basic Local Alignment Search Tool
EMBL	European molecular biology laboratory

Chapter I

Introduction

1. Background

Death is a profound motivator. 120 years ago, a healthy man was hospitalized with worsening symptoms of spreading pain and recurrent fever, yet he died with no definitive diagnosis. Following post mortem autopsy, a case of endocarditis was described along with vegetative growths on the valves of his heart, spleen, lungs, kidneys, and within blood vessels. Samples were taken and cultured using growth media, which revealed the presence of phenotypically identical bacteria. This bacterium first described as *Micrococcus zymogenes* was observed under the microscope as very small cocci arranged in pairs or chains, with a strong positive reaction to gram's stain, and a characteristic ability to acidify, coagulate, then liquefy milk. Inoculation of the cultured bacterium into mice and dogs induced symptoms of generalized infections throughout their bodies, as well as pronounced abscesses on key organs as mentioned above (MacCallum & Hastings, 1899). This prompted the scientists of that time to struggle in identifying, characterizing, and also treating infections caused by enterococci.

The aforementioned bacterium, currently identified as *Enterococcus faecalis*, has ever since established itself as a highly resilient organism capable of withstanding broad ranges of pH, temperature, antiseptics, and even nutrient starvation (Tiercelin & Jouhaud, 1899; Brodrick, 2016). *E. faecium* and *E. faecalis* have evolved in the past few decades to become highly potent opportunistic pathogens. They are leading causative organisms of multidrug resistant hospital acquired infections. They pose a

huge challenge to clinicians and public health officers as treatment options currently available are rapidly depleting (WHO, 2017), and their phylogenetic heterogeneity is proving to evolve fast enough to necessitate the need for constant improvement of genomic typing and surveillance methods (van Schaik, 2010; Leberton, 2013).

2. Taxonomy: Species classification and re-classification

Andrewes & Horder identified the organism as *Streptococcus faecalis.* "*Streptococcus"* to emphasize its morphological similarity to streptococci as chained cocci and strong reaction to gram's stain, and "*faecalis"* to also highlight the fact that this unique streptococcus had been isolated from the human gut. Little was known about enterococci and their pathogenic capabilities at that time, and these sporadic studies established the ground on which today's impressive knowledge base has grown. (Andrewes & Horder, 1906)

Previously, enterococci and streptococci were all classified under the same umbrella as there were no clear characteristics that would identify them as an independent genus or species with their unique traits. As the cases of deadly infections increased over time, there was the need for more organized classification of these destructive microbes. Sherman (1937) identified the key features that distinguish certain species of these cocci from other microorganisms of similar morphology. Sherman noted how enterococci boast a broad growth range of 10-45°C, are Gram-positive, and do not form spores. They can survive as obligate chemoorganotrophic fermenters, and even as facultative anaerobes. He later curated a scheme to differentiate between streptococci (including enterococci at that time) based on their hemolytic and proteolytic abilities. The scheme divided streptococci into *S. faecalis,*

S. faecalis var. *hemolyticus*, *S. faecalis* var *zymogenes*, and *S. faecalis* var *liquefaciens*.

DNA-rRNA hybridization studies finally provided the genetic basis upon which enterococci were classified as a unique genus (Schleifer & KilpperBalz, 1984). This led to the rise in the number of identified enterococcal species starting with the most noteworthy *E. faecalis* and *E. faecium*, to more than 20 identified species in the following decade, with the number of identified species currently being over 50 (Guzman Prieto, 2016; Bonacina, 2016).

Primitive typing methods based on biochemical assays, colony morphology, or even microscopic features of the cells could not differentiate between the diverse populations of enterococci. Genetic assays based on housekeeping genes, in particular the 16S and 23S rRNA coding regions were used as alternative typing methods to finally understand where these bacteria stand in the tree of life. In fact, 16S rRNA sequence analysis showed that enterococci and streptococci were actually more distant than previously thought, and enterococci were closer to teragenococci and vagococci than streptococci (Facklam, 2002).

As typing methods evolved and new enterococcal species were discovered, more enterococcal species rapidly were identified. This challenged the available typing approaches and pushed for new identification methods with higher resolution. Single cell protein analysis and 16S rRNA analysis once seen as the gold standards in the late 90s took a blow when it came to differentiating *E. gallinarum* and *E. casseliflavus* as these species had an almost identical (99.9%) 16S rRNA coding region (Patel et al, 1998). Consequently, alternative methods were developed and adopted in search for higher typing resolution, such as: Randomly Amplified

Polymorphic DNA (RAPD) analysis (Monstein, tiveljung et al 2000, Quednau, ahrne, et al 1998), sequencing of 23S rRNA spacer regions (Naimi et al, 1999) and sequencing of highly conserved protein coding regions *sodA* (Poyart et al, 2000), *efaA* (Singh et al, 1998), *cpn60* (Goh et al, 2000), and *ddl* (Ozawa et al, 2000).

3. Enterococci in Food

Due to their ubiquitous presence in nature, enterococci are frequent colonizers of food. Enterococci are important members of the lactic acid bacteria, as they are important fermenters of cheese and some meat products, but from a public health standpoint, their presence in food is of concern when addressing cases of clinical relevance (high levels of resistance).

Resistance to antimicrobials and sanitizers used during food processing, as well as tolerance of the harsh conditions of cooking or pasteurization enable enterococci to survive and show a marked presence in food such as in Irish Baylough cheese (Gelsomino et al, 2004), Appenzeller raw milk cheese (Templer et al, 2006), Pecorino Abruzzese cheese (Serio et al, 2007), and even dry fermented sausage (Ribeiro et al, 2011) . Most of the studies based on isolates from food sources highlighted antimicrobial resistance genes and altered phenotypes enabling these bacteria to reach the same levels of robustness in survival mechanisms as that of clones circulating in the hospital environment. Ingestion of these food products gives enterococci access to the GI track of humans and can lead to serious infections if they cross the lumen of the intestine (Giraffa, 2002; Franz et al, 2003).

4. Enterococci as Gut Commensals

Both *E. faecium* and *E. faecalis* are natural inhabitants of the human gut microbiome. Comprising only around 1% of the bacterial population of the intestine of healthy

adults, antibiotic resistant enterococcal sub-populations rapidly dominate and colonize the GI track after its depletion from healthy commensal bacteria following broad spectrum antimicrobial therapy (Donskey et al, 2000; Ubeda et al, 2010). This switch from commensal to pathogen was further studied by injecting *E. faecalis* into the midgut of a model organism *Manducta sexta* that showed how enterococci could survive in the harsh digestive environments of the gut, and could easily pass into the blood to cause sepsis if a pore forming toxin damages the epithelium (Mason et al, 2011).

With the limited antimicrobials available to treat these emerging super-resistant clones of enterococci, efforts have been directed towards using probiotics in attempts to reduce overgrowth and prevent the spread and invasion of the bloodstream through the epithelial lining of the GI track (Crouzet et al, 2015).

5. Enterococci as Hospital Associated Pathogens

In the clinical setting, enterococci are encountered in soft tissue inflammations, more specifically endocarditis (MacCallum & Hastings, 1899), urinary tract infections, and bacteremia. Enterococci are linked to over 1 in 10 Urinary tract infections (UTI) that are catheter associated (Weiner et al, 2016), causing a huge burden on healthcare professionals with very limited treatment options. It is also a major public health concern due to its resilience and genomic plasticity enabling it to spread through hospital wards (Boyce et al, 1995; Caballero et al, 2017).

Nosocomial outbreak officers have been in a continuous battle against the spread of enterococci. Bacterial identification methods and typing schemes are continuously

being improved to gain better insights on the propagation of these pathogens in the hospital environment, eventually causing harm to the community (Raven et al, 2016).

Multilocus sequence typing was coined as a successful typing scheme to differentiate between environmental and more pathogenic hospital associated enterococci (Homan et al, 2002). This highlighted the hospital associated clonal complex 17 (CC17) (Willems et al, 2005) that includes *E. faecium* strains that are generally more virulent and have more mobile genetic elements and resistance determinants (Willems et al, 2005; leavis et al, 2006; kim & Marco, 2014).

However, from an evolutionary standpoint, the high genomic plasticity of *E. faecalis* and *E. faecium* governed by mobile genetic elements and high rates of gene recombinations under selective pressure gave it a competitive edge in survival (Turner et al, 2007). This, however, hinders the straightforward phylogenetic classification and outbreak monitoring processes as this makes it difficult on healthcare professionals to differentiate between highly similar circulating clones.

Some opt to use the classical approach of sequence typing based on the seven housekeeping genes and show clear cut lines of difference between bacterial populations (Pinholt et al, 2015). However, using whole-genome sequence data it's now possible to target more than a thousand loci, such as the core genome multi locus sequence typing (cgMLST) scheme, which can differentiate between isolates (Santona et al, 2016), providing better resolution and easier segregation of enterococci that were previously deemed nearly identical.

6. Epidemic Strains

In the past decade, pathogenic enterococcal clones were clustered under the global clonal complex 17 (CC17), which includes sequence types ST-17, ST-18, and ST-78

(Willems et al, 2012). However, with the increasing pressure on enterococcal clones, a pronounced heterogeneity arising from recombinational events and focused at specific loci was observed.

A noteworthy study relevant to our region is the detection of a novel sequence type ST910 of vancomycin resistant enterococci isolated and characterized in Tunisia, showing how independent clones of very high potential threat are emerging worldwide (Ben said et al, 2016).

The already pronounced heterogeneity of vancomycin resistant enterococcal (VRE) clones is nowhere short of reaching its end, as huge studies in developed countries are conducting bioinformatic analyses on whole genome sequence data to understand the population structure of these VRE clones. Examples of large scale projects include characterization of the newly emerging ST192 clone with VanB type resistance (Bender et al, 2016), the SENTRY antimicrobial surveillance program in the United States (Deshpande, 2015), pilot studies to uncover novel sequence types and uniquely evolved genomes (Tavares, 2019), and huge efforts in the United Kingdom (Reuter, 2013; Gouliouris, 2018) and Japan (Kuroda, 2018) aimed at surveillance, outbreak prevention, and phylogenetic analyses to uncover evolutionary relationships of clinically relevant enterococci

In Australia, a novel clone identified as ST796 emerged as a result of recombination events that evolved from the previously prevalent ST555 clone circulating in the region. *E. faecium* also showed the ability to acquire integrative phage sequences, cryptic genomic islands, and mosaic plasmid sequences, as well as the two famed transposons that made it resistant to vancomycin (Tn1549) and tetracycline (Tn916) (Buultjens et al, 2016). Vancomycin resistant enterococci (VRE) of extremely diverse

clones have also been identified in Iran (Fatholahzadeh et al, 2006), and a study in Saudi Arabia highlighted two new Tn1546-like variants of the vancomycin resistance cassette (Khan et al, 2008).

7. Antimicrobial Resistance

Some of the earliest incidences of antimicrobial resistance were seen in cases of endocarditis where some patients did not respond positively to treatment with penicillin. In the 1950s, there was still nothing separating enterococci from streptococci, even though this major difference in resistance phenotype was apparent from the onset of antibiotic use to treat infections.

Streptococcal infections were readily treated with penicillin at first, while enterococci falsely identified as streptococci were resistant to the treatment since they produce penicillin binding proteins of low affinity on their cell walls (Bourgeois-Nicolaos, 2014). This prevents the efficient action of penicillin G, ampicillin, and even cephalosporins on the cell walls of enterococci. Apart from these intrinsic resistance determinants, enterococci are extremely adept to exchanging mobile genetic elements and plasmids to readily gain coding sequences of DNA that broaden their resistance patterns (Arias, 2012).

8. Vancomycin Resistance

Vancomycin is a glycopeptide antibiotic used to treat certain infections caused by Gram-positive cocci by targeting the D-Alanine-D-Alanine (D-Ala-D-Ala) chain of the peptidoglycan layer of the cell wall. Enterococci have accordingly evolved to synthesize variants that are more resistant to glycopeptides, specially vancomycin. This is achieved through inserting D-Alanine-D-Serine (D-Ala-D-Ser) or D-Alanine-D-Lactate (D-Ala-D-Lac) chains in the peptidoglycan layer instead of the traditional

D-Ala-D-Ala linkage. The variants, due to the difference in spatial arrangements of these polypeptide chains, have differential levels of resistance to vancomycin. Mutants with a cell wall chained with D-Ala-D-Ser show around seven times less binding affinity, while D-Ala-D-Lac mutants exhibit extremely high levels of resistance with binding affinity to vancomycin of less than 1000 fold compared to the wildtype (Mendez-Alvarez, 2000, Arthur, 1993).

Vancomycin is one of the last resort antimicrobials available to treat enterococcal infections, and vancomycin resistance is a major field of interest. Resistance to vancomycin is linked to variants of the *van* operon, denoted as *vanA, vanB, vanC, vanD, vanE, vanG, vanL, vanM* and *vanN.* Each of these operons differs from the others in the arrangement of the *van* genes that make up the resistance cassette. Depending on the sequential arrangement of the *van* genes, these operons code for varying levels of phenotypic resistance in the enterococci harboring them (sifaoui 2001, Courvalin et al, 2006).

Clinically relevant Vamcomycin Resistant Enterococci (VRE) clones are usually of *vanA* or *vanB* genotype, while other variants are found in food and environmental vancomycin resistant enterococcal isolates. Both *vanA* and *vanB* operons comprise a two-component system having the *vanS*, sensory kinase, and the *vanR*, the response regulator. These two genes promote the expression of downstream genetic elements ultimately reducing the binding of vancomycin to the peptidoglycan pentapeptide.

The *vanA* operon is usually carried on a Tn1546 transposon or a Tn1546-like transposon that is slightly different due to single nucleotide variants. It was also detected on structural variants of the same Tn1546 transposon. These structural variants including insertions, deletions, and some inversions are induced by insertion

sequences or represent evolutionary footprints of genetic change observed as the transposon propagates between isolates. *vanA* positive isolates, and upon activation by the *vanS/R* two component system induces the expression of *vanH*, which codes for a dehydrogenase that catalyzes the conversion of pyruvate to lactate, and *vanA* encoding a ligase that links Alanine to Lactate forming the vancomycin resistant variant D-Ala-D-Lac dipeptide (Hegstad et al 2010, Werner et al, 2011).

The *vanB* operon on the other hand, encodes homologs of the *vanR, vanS, vanH, vanX,* and *vanY* genes the expression of which will also alter the peptidoglycan layer interfering with the binding of vancomycin. However, *vanB* additionally codes for an uncharacterized protein *vanW* and the role it plays is yet to be uncovered. *vanB* is usually carried on Tn1547 (or Tn1547-like), Tn1549 (or Tn1549-like), Tn5382 (or Tn5382-like) transposons that are typically of chromosomal origin and not usually harbored on plasmids (Launay et al, 2006, Lopez et al, 2009).

9. Whole-Genome Sequencing of VRE

With the rapid advances in whole-genome sequencing (WGS) technologies and the steep decline in cost, a fully annotated whole-genome sequence of a clinical bacterial isolate can be available in roughly 3 days for just under 100$ on an Illumina Miseq whole-genome sequencer. Initially, whole-genomes of bacteria were analyzed to annotate functional genes and to identify genetic variants that encode resistance or virulence determinants that are of clinical relevance. Gradually bioinformatic analysis software tailored for prokaryotic genomes came into existence. The simple one-dimensional scripts that used to perform Basic Local Alignment (BLAST) that indicate presence/absence of a gene in a sequenced fragment of DNA have evolved throughout this past decade into curated pipelines of bacterial genome analysis. The

open source nature of this software, along with an extremely supportive community of both developers and users have cultivated huge advocacy for end-to-end bioinformatics-based bacterial genome analysis in recent years. The EMBL and Sanger institutes have put in their fair shares of efforts to both develop and maintain prokaryotic genome analysis tools of the highest caliber. Additionally, due to the open source, peer reviewed, and sometimes community-developed nature of the analysis pipelines, independent researchers have introduced alternative analysis methods that tackle certain areas of research that require more specialized tools on top of the more general "one size fits all" software.

An outbreak linked to VRE in Denmark was analyzed using bioinformatic analysis of whole-genome sequences of 495 isolates in order to better understand the evolutionary dynamics and phylogenetic relatedness of circulating VRE clones in the region. Although all clones carried the *vanA* operon on a Tn 1546 transposon, there was significant variability in the Tn1546-like fragments with five novel transposons being identified, as well as the establishment of phylogenetic clades that segregated the isolates in question into clearly differentiable clones circulating in major hospitals in the country (Pinholt et al, 2016). Another study in Australia (Leong, 2018) also analyzed the whole-genomes of VRE clones circulating in a hospital, determining as a result the transmission routes of VRE between hospital wards. Such high discriminatory power to analyze the phylogenetics of almost identical isolates is only possible through the utilization of modern bioinformatic analyses on whole-genome sequences of enterococcal isolates in question

Enterococci are evolving at a rate higher than our ability to produce antimicrobials and maintain rigorous outbreak surveillance schemes. There is the need for faster, more efficient, and highly discriminatory methods to manage this pathogen. To our

knowledge, this is the first study of its kind in profiling a sample of the clinically relevant enterococcal clones circulating in Lebanon. We aimed to uncover the genetic basis of the observed phenotypic resistance to antimicrobials used as the standard of care, as well as to highlight the heterogeneity of enterococcal clones using whole genome sequence based phylogenetic analyses.

Chapter II

Materials and Methods

1. **Ethical approval**

Ethical approval was not required as the clinical isolates were collected and stored as part of routine clinical care. Clinical isolates and patients' record/information were anonymized and de-identified prior to analysis.

2. **Bacterial Isolates**

A total of 21 isolates collected in 2017 were provided by the American University of Beirut Medical Center (AUBMC) and were designated as EF1-EF21.

66.7% (n=14) were males and 33.3% (n=7) were females (Figure 1).

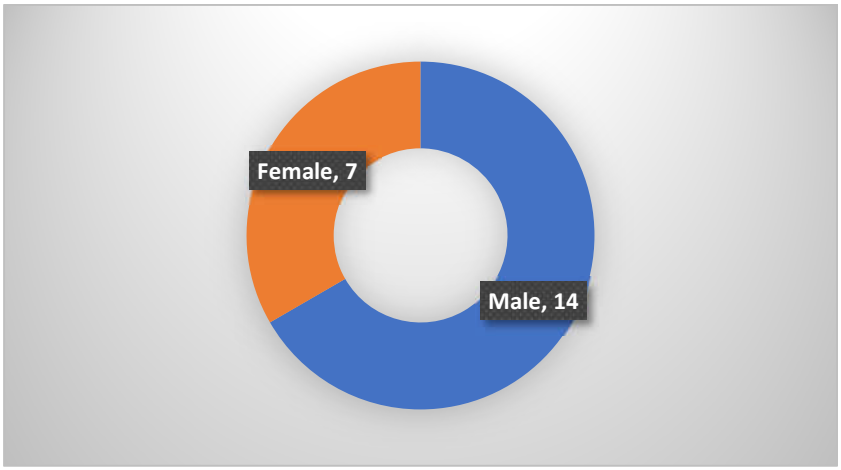


Figure 1. Diagram showing samples distribution between both sexes.

Patients' ages ranged from 30 to 91 with a median age of 65 (Figure 2).

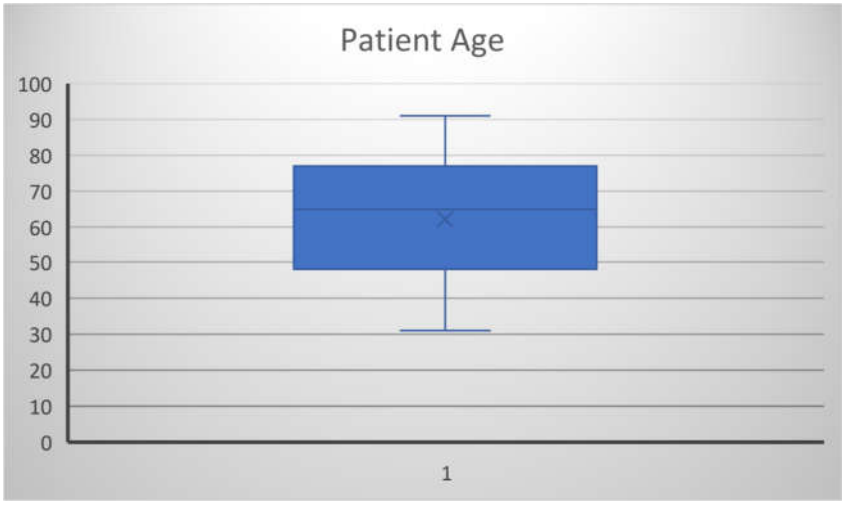


Figure 2. Box and Whisker plot showing the distribution of patients' ages

Isolates were collected from abdominal fluid (n=1), blood(n=7), deep tracheal aspirate (n=3), tissue(n=1), urine(n=5) and wound (n=4) (Figure 3).

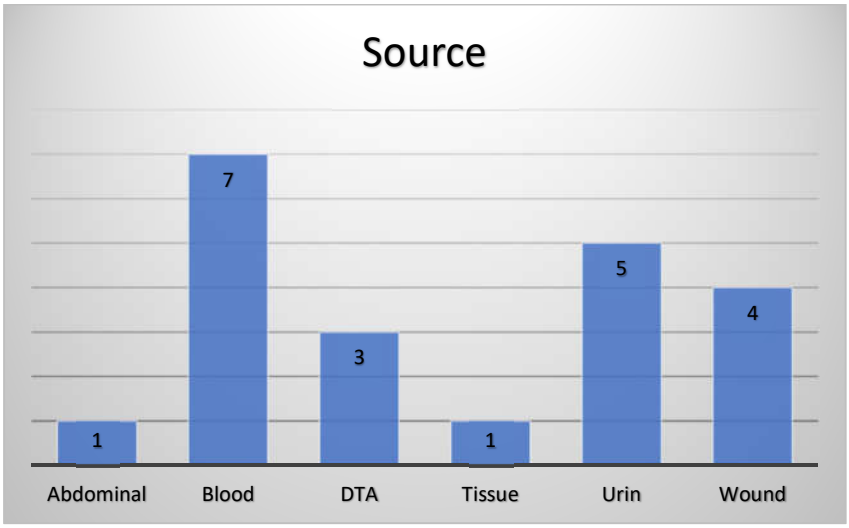


Figure 3. Bar chart showing the collection sites of the various isolates.

3. **MALDI-TOF**

Initial species identification was performed by the Matrix-Assisted Laser Desorption/Ionization Time of Flight (MALDI-TOF) system (Bruker Daltonik, GmbH, Bremen, Germany) following the manufacturer's instructions.

4. **Antimicrobial susceptibility testing**

According to current standards of care based on phenotypic microbiological laboratory analysis, all isolates were profiled based on the sex and age of the patient, source of isolation, as well as phenotypic susceptibility to the antimicrobials used to commonly treat enterococcal infections; Ampicillin, Erythromycin, Teicoplanin, and Vancomycin. Antimicrobial susceptibility testing was performed using the Kirby-Bower disk diffusion assay for Ampicillin and Erythromycin. Vancomycin and Teicoplanin minimal inhibitory concentrations (MICs) were obtained using the E-test. Results were interpreted based on the CLSI guidelines (Weinstein, M. P. 2019).

5. **DNA Extraction**

Genomic DNA (gDNA) was extracted using Nucleospin Tissue kit (Macherey Nagel) according to the manufacturer's guidelines for gram positive bacteria.

6. **Whole-Genome Sequencing**

gDNA was used as input for library preparation using the Illumina Nextera XT DNA library preparation kit (Illumina), and 1 μg of DNA was used as input for library preparation. The gDNA was subjected to end-repair, A-tailing, ligation of adaptors including sample-specific barcodes as recommended in the manufacturer's protocol. The resulting library was quantified using a Qubit 2.0 fluorometer. The

library was sequenced on an Illumina MiSeq with paired-end 500 cycles protocol to produce paired end reads of length 250 bp.

7. **Quality Control and Assembly**

Quality control of the whole genome sequence reads was performed using FastQC (Andrews, 2010). Low quality reads were demultiplexed and trimmed of adapter fragments or low quality (Minimum phred Q30) bases using Trimmomatic (Bolger, 2014). Assembly was performed *De Novo* using SPAdes V 3.6 (Bankevich, 2012).

8. **Annotation**

Whole genomes were annotated using Prokka (Seemann, 2014), using relevant settings to specify the following: -kingdom bacteria -genus Enterococci – gram pos -locustag EF. Protein coding sequence alignment identity of 95% was selected for higher accuracy of automated annotation. Annotated genomes in GFF3 format were used in downstream analyses of genomic islands and pan genome analysis.

9. **MLST**

Multilocus sequence typing (MLST) was performed based on seven housekeeping genes for *E. faecalis* and *E. faecium* using MLST 1.8 database available on www.genomicepidemiology.org/.

10. **Pan Genome Analysis**

    A. Pan Genome of Coding Regions

The pan genomes of *E. faecalis* and *E. faecium* coding regions were generated using the Roary pipeline (Page, 2015) on the annotated genome files in GFF3 format.

Relevant settings were used to specify the following criteria (Table 1)

Table 1. Commands used to run the Roary pipeline for pan genome analysis of the coding regions and their associated functions

Command	Function
-e	To generate a multiFASTA alignment of core genes using PRANK (Löytynoja, 2014)
-n	to generate a fast core gene alignment with MAFFT (Katoh, 2013)
-r	to generate R statistical plots using ggplot2 (Wickham, 2016)
-v	verbose output for troubleshooting
-s	keeps paralogs together – important downstream in Piggy

B. Pan Genome of Non-coding Intergenic (IGR) regions

The pan genomic analysis of non-coding intergenic regions was performed using the Piggy pipeline (Thorpe, 2018). This approach emulates the previous step of the analysis but in contrast it generates a presence/absence matrix based on "switched" IGR regions as well as the downstream genes. The gene presence/absence matrices with their phylogenetic trees were visualized using Phandango (Hadfield, 2017).

11. **Core Genome Single Nucleotide Polymorphism Analysis**

The reference genomes of *E faecalis* V583 (NC_004668) and *E faecium* DO (NC_017960) were used to call single nucleotide polymorphisms in the core genomes shared by all isolates of each species. The Snippy pipeline (Seemann,

2015) was used to align whole genome sequence reads of each species to each other and to call for variants altering a shared locus across all genomes.

Variants of interest, as well as run parameters for FREEBAYES are noted in Table 2

Table 2. Variants of interest in core genome single nucleotide polymorphism analysis (cgSNP) and FREEBAYES parameters used for the discovery of polymorphisms and genetic variants

Abbreviation	Explanation
SNP	Single nucleotide polymorphism
MNP	Multiple nucleotide polymorphisms
INS	Insertion
DEL	Deletion
CPX	Complex (SNP and/or MNP)
FREEBAYES - MINCOV 15	minimum number of reads covering a site to be considered
FREEBAYES - MINQUAL 100	minimum VCF variant call quality

12. **Core Genome Phylogenetic Analysis**

The core genome alignment was used to generate a maximum likelihood phylogenetic tree with RAxML V 8.0 (Stamatakis, 2014). The GTR-GAMMA model of heterogeneity was adopted, and a phylogenetic tree was generated based on 1000 bootstrap iterations. The phylogenetic tree was visualized using the Interactive Tree of Life (ITOL) (Letunic, 2006).

13. **Recombination Analysis**

Recombination analysis was performed using the Gubbins pipeline (Croucher, 2014) following recommended options. Ancestral sequences were reconstructed using both

RAxML (Stamatakis, 2014) following the GTR-CAT model with 500 iterations (for accuracy) and FastML (Ashkenazy, 2012) (for speed) giving identical results. Recombination matrix and associated phylogenetic tree were visualized using Phandango (Hadfield, 2017).

14. **Resistance Genes and Comparative Genome Analysis**

Resistance genes were inferred using Resfinder 3.0 (Zankari, 2012). Average nucleotide identity (wgANI) analyses were carried out using the web tool provided by Kostas Lab (http://enve-omics.ce.gatech.edu/ani/).

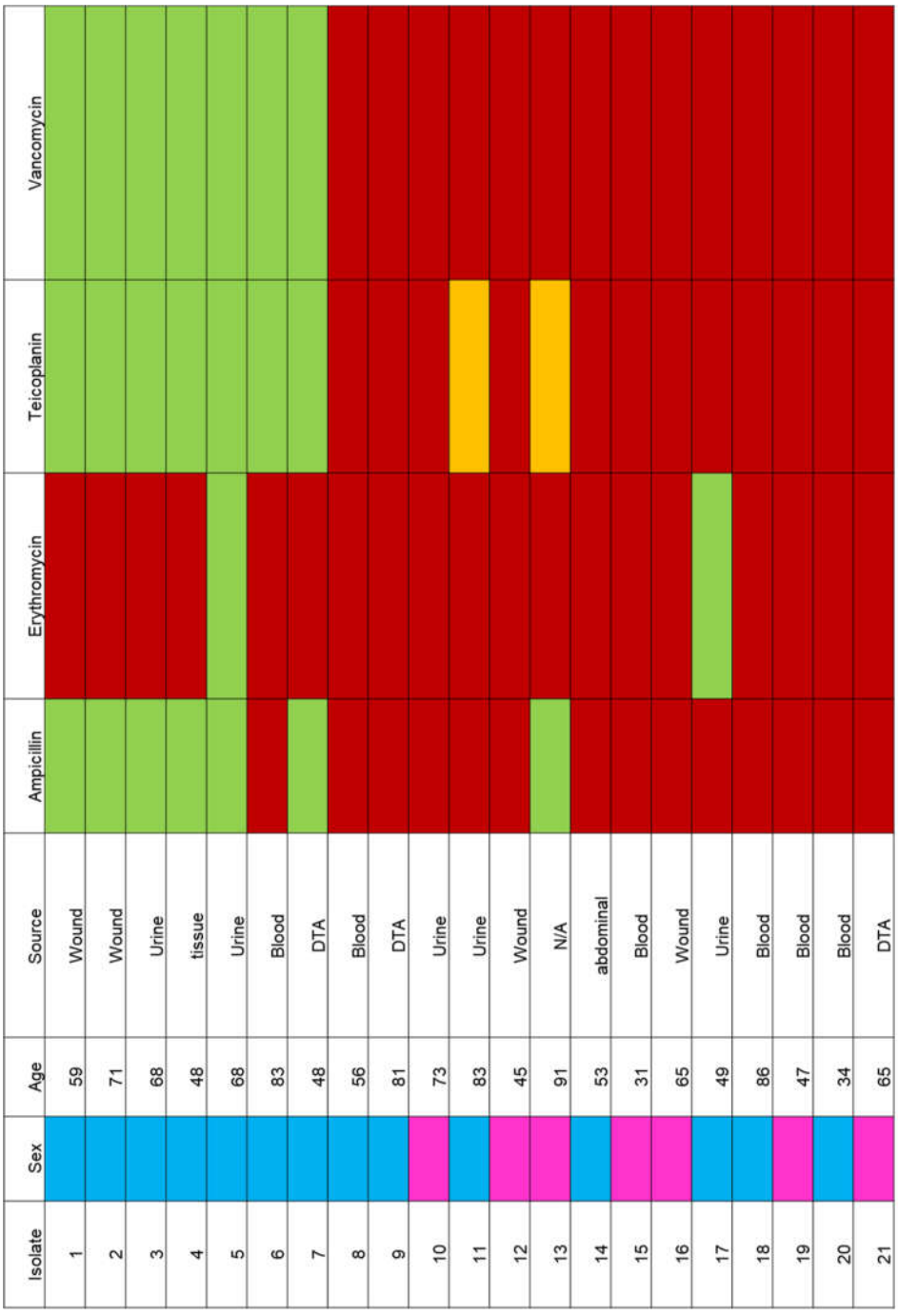
Chapter III

Results

1. Phenotypic antimicrobial susceptibility profiling

Isolates were differentially resistant to the tested antimicrobials with no clear-cut patterns of linked inheritance observed (Table 3). In total, 14/21 (66%) of the isolates were resistant to Ampicillin, 19/21 (91%) were resistant to Erythromycin, and 14/21 (66%) of the isolates were resistant to both Teicoplanin and Vancomycin. Notably, only two isolates (EF11, EF13) showed intermediate resistance to Teicoplanin. Isolates resistant to Ampicillin were also resistant to Erythromycin, except for EF17, which was resistant to Ampicillin but not to Erythromycin. The sex and age of the patients, and even source of isolation of the bacteria had no correlation with the ST-types or the phenotypic resistance profiles of the isolates.

Table 3. Shows the patient's sex (Blue – Male, Pink – Female), age, and source of isolation. Phenotypic resistance to Ampicillin, Erythromycin, Teicoplanin, and Vancomycin are annotated as Sensitive – Green, Resistant – Red, Intermediate – Orange .



2. *in Silico* MLST and Resistance profiling

in Silico multilocus sequence typing was performed using the whole-genome sequences of the isolates, classifying them as either *E. faecalis* or *E. faecium* of known ST-types (Table 4). 8/21 (38%) of the isolates belonged to ST-80, 5/21 (24%) of the isolates belonged to ST-203, while others uniquely matched to ST-28, ST-40, ST-55, ST-117, ST-376, and ST-480. Notably, isolate EF13 only matched to 5/7 of the alleles of ST-224 of the *E. faecium* sequence typing scheme giving inconclusive results for its sequence type.

in Silico resistance profiling based on known resistance determinants in the whole-genomes of the isolates (Table 5) showed that all isolates except EF5 (20/21, 95%) were resistant to Aminoglycosides, Macrolides, Lyncosamide, and Streptogramin B. Genetic determinants encoding resistance to Phenicol (9/21, 43%) , Tetracycline (15/21, 71%), Trimethoprim (8/21, 38%), Oxazolidinone (2/21, 9%), and most importantly Glycopeptides, of which Vancomycin/Teicoplanin (14/21, 66%), were identified. Resistance to Aminoglycosides was encoded by *ant(6)-Ia, aph(3')-III,* and *aac(6')-aph(2'').* The presence of *lsa(A), erm(B), Msr(c), erm(T),* and *Inu(B)* indicated resistance to Macrolides, Lincosamide, and Streptogramin B. Tetracycline resistance was identified by the presence of the *tet (M/L/S)* genes, while resistances to Trimethoprim and Oxazolidinone were encoded by the *dfrG* and *optrA* genes, respectively.

Resistance to Vancomycin and Teicoplanin was encoded by the vancomycin resistance cassette of VanA-type, encoded by *vanR, vanH, vanA, vanX, vanS, vanZ, and vanY.* Only one of the vancomycin resistant isolates, namely EF15 was negative for *vanZ* and *vanY*.

Table 4. MLST profiles and *in silico* predicted resistance patterns of the isolates based on whole-genome sequences. Red – Resistant; Blue – Sensitive

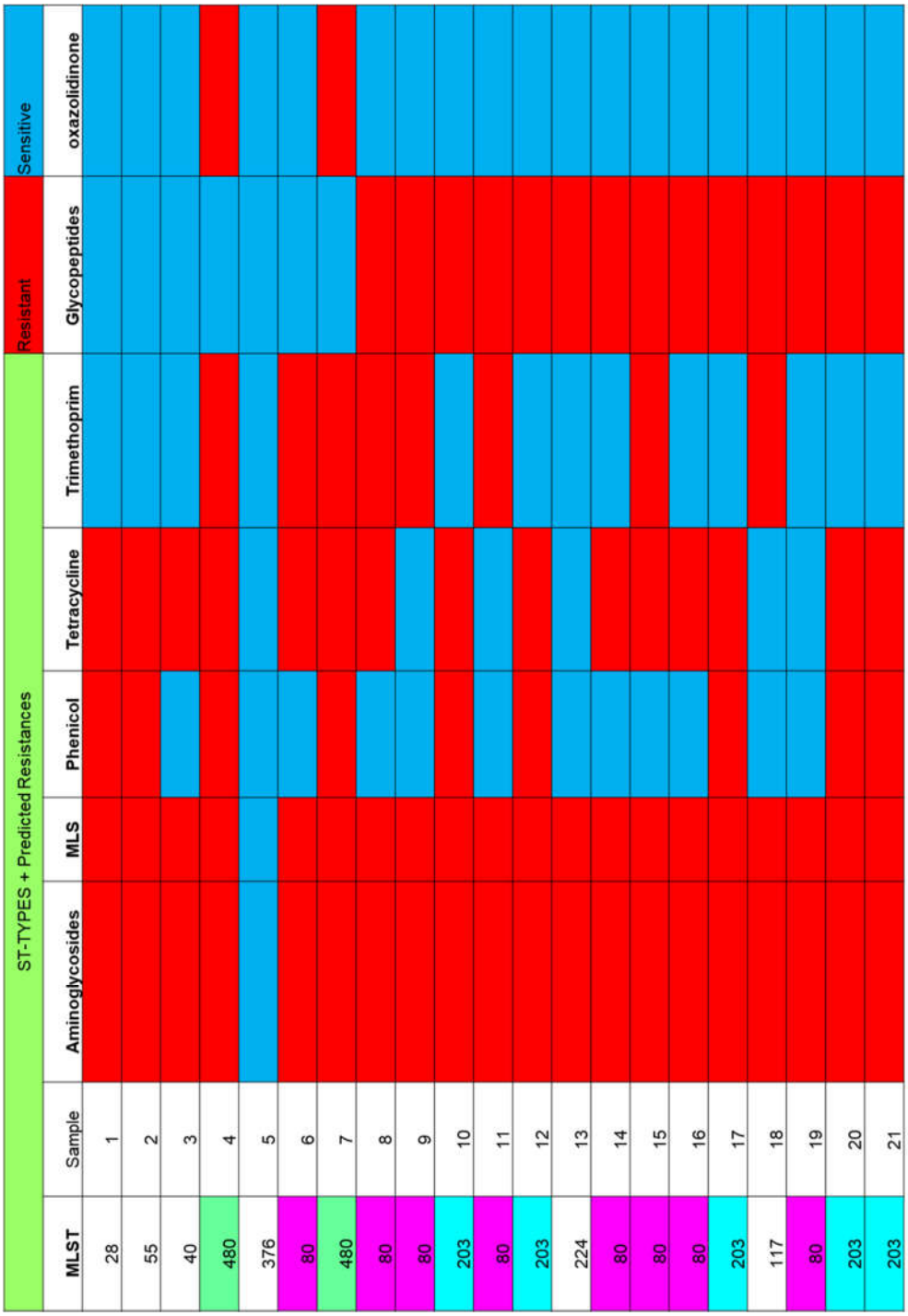
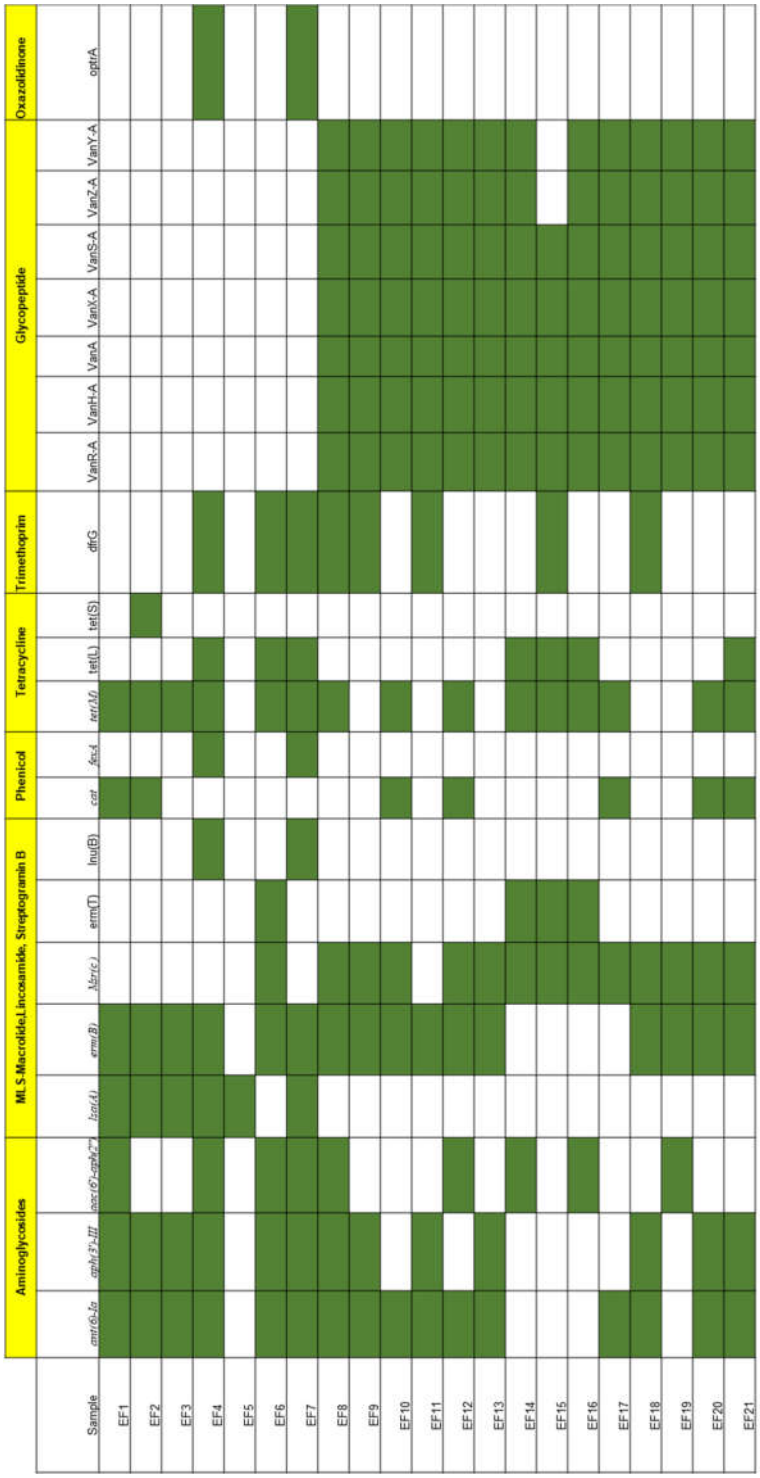


Table 5. Resistance determinants mediating resistance to different antimicrobial agents based on *in silico* gene alignments. Green – Presence; White – Absence



3. Whole genome single nucleotide polymorphism (wgSNP) based phylogeny

   Phylogenetic analysis of the isolates based on whole-genome single nucleotide polymorphisms was initially performed on the assemblies (.fasta) of the genomes in order to generate a crude phylogenetic overview of the isolates. A similarity matrix (Table 6) was generated based on the single nucleotide variants between pairs of the isolates. Pairwise SNP distances were around 30 or less for isolates of the same clade, and more than 4000 SNPs were observed between isolates of different clades. Isolates EF 5, EF 4, EF 7, EF 1, EF 3, and EF 2 comprise clade A, having only around 30 SNPs between them, whereas clade B composed of isolates EF 12, EF 20, EF 17, EF 10, EF 11, EF 9, EF 19, EF 6, EF 15, EF 14, and EF 16 was different than clade A by over 4000 SNPs. Isolates EF 13 and EF 18 were too divergent to be classified as part of either clade.

   Two phenotypically different isolates, namely the vancomycin resistant isolate EF15 and the vancomycin sensitive isolate EF6 showed identical SNP profiles in their assemblies, even though these isolates had diverse resistance determinants in their genomes. These two isolates had an average nucleotide identity (ANI) of 99.91%, and were of the same sequence type ST-80. The alignment of the assemblies was used to generate phylogenetic trees to visualize the genomic heterogeneity and evaluate isolate relatedness. The phylogenetic tree clustered *E. faecalis* and *E. faecium* isolates separately, with a few isolates not identified as belonging to either cluster A or cluster B, and isolates EF 14 and EF 18 clustered on their own. The Multidimensional scaling map (MDS) demonstrated by the circular phylogenetic tree (Fig 5) also demonstrated the tight clustering of most of the isolates into two major clades, with a few divergent genomes situated far from the central cluster. Cluster B was more central, while isolates of cluster A were situated in the

periphery. Isolate EF 13 was observed to be in between the two major clades described, while isolate EF 18 appeared more ancestral than the central clade and clustered on its own.

Table 6. Matrix of pairwise genetic distances of the isolates based on the alignment of whole-genome assemblies, (.fasta)

wgSNP Distance Matrix

Samples	18	19	3	12	6	20	15	8	11	17	4	14	16	1	2	9	7	5	10	13
18	-	24	4342	14	25	14	25	23	19	14	4338	22	21	4351	4340	28	4338	4341	14	299
19	24	-	4339	32	5	32	5	13	7	32	4335	6	5	4348	4337	16	4335	4338	32	307
3	4342	4339	-	4351	4337	4351	4337	4342	4341	4351	50	4339	4338	48	20	4350	50	60	4351	4330
12	14	32	4351	-	33	0	33	31	27	0	4347	30	29	4360	4349	36	4347	4350	0	295
6	25	5	4337	33	-	33	0	10	8	33	4333	11	10	4346	4335	17	4333	4336	33	306
20	14	32	4351	0	33	-	33	31	27	0	4347	30	29	4360	4349	36	4347	4350	0	295
15	25	5	4337	33	0	33	-	10	8	33	4333	11	10	4346	4335	17	4333	4336	33	306
8	23	13	4342	31	10	31	10	-	8	31	4338	11	10	4351	4340	17	4338	4341	31	306
11	19	7	4341	27	8	27	8	8	-	27	4337	5	4	4350	4339	9	4337	4340	27	304
17	14	32	4351	0	33	0	33	31	27	-	4347	30	29	4360	4349	36	4347	4350	0	295
4	4338	4335	50	4347	4333	4347	4333	4338	4337	4347	-	4335	4334	60	48	4346	0	54	4347	4327
14	22	6	4339	30	11	30	11	11	5	30	4335	-	1	4348	4337	13	4335	4338	30	305
16	21	5	4338	29	10	29	10	10	4	29	4334	1	-	4347	4336	13	4334	4337	29	304
1	4351	4348	48	4360	4346	4360	4346	4351	4350	4360	60	4348	4347	-	52	4359	60	76	4360	4342
2	4340	4337	20	4349	4335	4349	4335	4340	4339	4349	48	4337	4336	52	-	4348	48	54	4349	4328
9	28	16	4350	36	17	36	17	17	9	36	4346	13	13	4359	4348	-	4346	4349	36	313
7	4338	4335	50	4347	4333	4347	4333	4338	4337	4347	0	4335	4334	60	48	4346	-	54	4347	4327
5	4341	4338	60	4350	4336	4350	4336	4341	4340	4350	54	4338	4337	76	54	4349	54	-	4350	4331
10	14	32	4351	0	33	0	33	31	27	0	4347	30	29	4360	4349	36	4347	4350	-	295
13	299	307	4330	295	306	295	306	306	304	295	4327	305	304	4342	4328	313	4327	4331	295	-

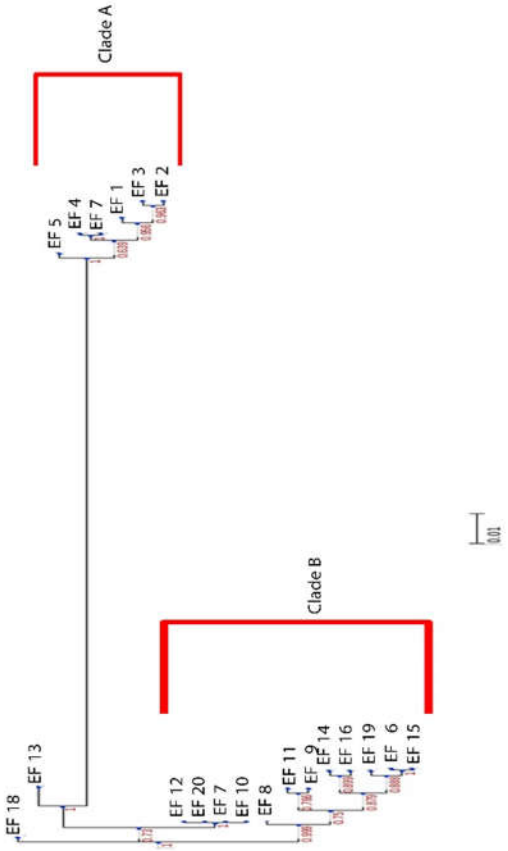


Figure 4. Maximum Likelihood (ML) phylogenetic tree based on whole-genome single nucleotide polymorphisms (wgSNP) of the aligned whole genome assemblies.

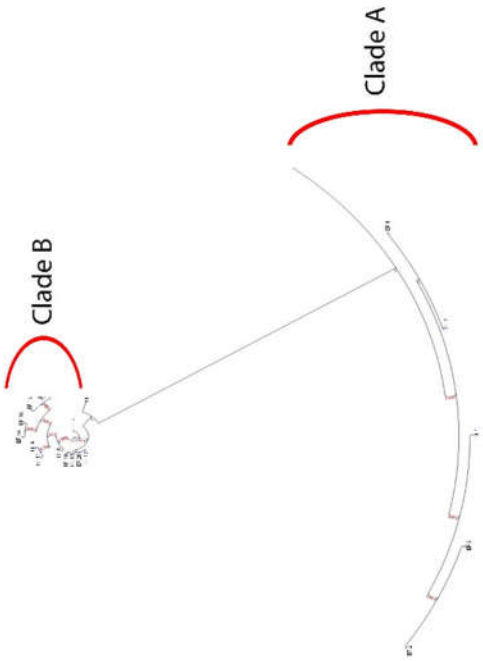


Figure 5. Circular maximum likelihood phylogenetic tree based on whole-genome single nucleotide polymorphisms (wgSNPs) of the aligned whole genome assemblies. Central clade (Clade B) and peripheral clade (Clade A) are observed along with non-clustering isolates EF 13 and EF 18 between them.

4. Pan Genome Analysis

All pan genome analyses were carried out separately for *E. faecium* and *E. faecalis* isolates; the isolates were split and analyzed separately to increase the intergenomic discriminatory power. Due to the large number of genes unique to each isolate, apparent from the high number of genes not part of the core genome, it was not feasible to highlight and discuss each gene individually. Thus, we selected genes uniquely found in some isolates but not all and that encoded proteins of clinical relevance.

A. Pan genomes of coding regions

1) *E. faecalis*

Pan genome analysis of all coding regions of *E. faecalis* identified a total of 3846 protein coding sequences found collectively in the isolates. Based on their prevalence in the genomes, the genes were identified as part of the core genome if they were present in all isolates. 2263 protein coding sequences constituted the core genome of *E. faecalis*. Alternatively, genes not present in all isolates but shared by most (prevalence of 15-95%) were designated as Shell genes with 1583 protein coding sequences belonging to this category in the generated *E. faecalis* pan genome (Figure 6).

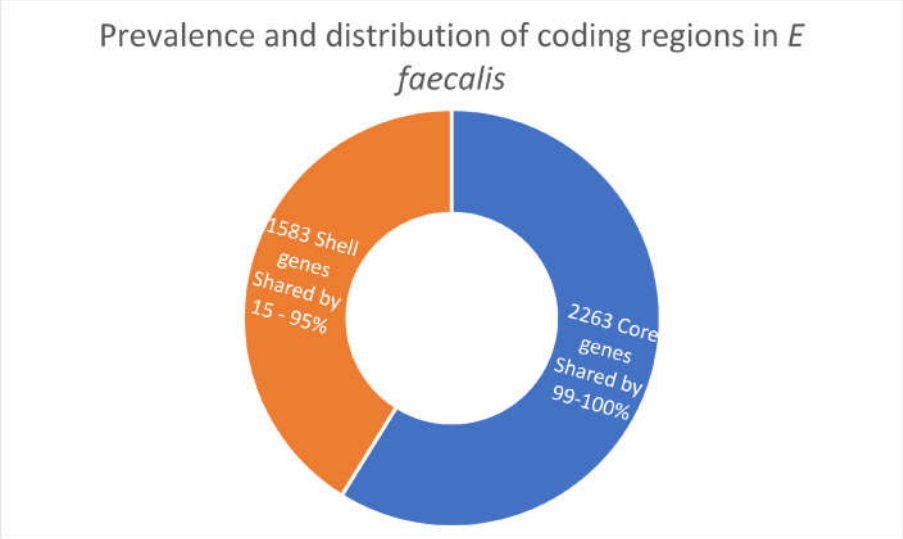


Figure 6. Circular diagram showing the prevalence and distribution of coding regions in *E. faecalis* isolates of our study

Figure 7 illustrates the relatedness of the isolates based on shared protein coding sequences. Each blue block represents a gene, and isolates are grouped based on the number of shared/unique genes as a phylogenetic tree on the left side of the matrix.

Isolates EF4 and EF7 shared the multidrug resistance protein (Stp), and a putative transposon Tn552 DNA-invertase. Isolate EF2 uniquely encodes the cold shock protein 2 (cspL), while EF3 uniquely encodes a putative prophage phiRv2 integrase, a DNA translocase (SftA), and a plasmid recombination enzyme (Pre). Isolate EF5 encodes for an accessory gene regulator A (AgrA), an antiseptic resistance protein (qacA), and a ferrochelatase (hemH). Interestingly, Isolate EF1 contained a number of genes coding for proteins that increase fitness in harsh environments and selective pressure such as tyrosine recombinase (XerC), chromosome partitioning ATPase (Soj), putative peptide export permease protein (YydJ), macrolide export ATP-binding permease (MacB), a few transposon associated DNA invertases, lipid flippase (MurJ , and proteins that regulate stress (gls24) and protect DNA during starvation.

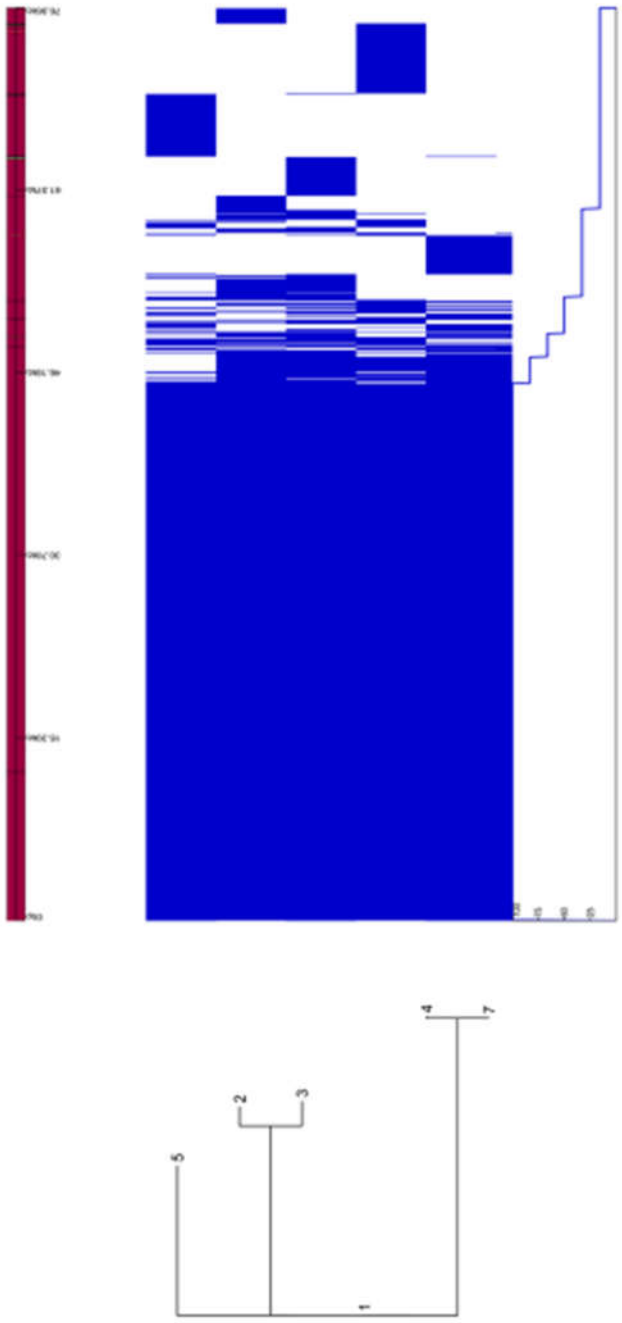


Figure 7. Presence/Absence matrix with its associated maximum likelihood phylogenetic tree based on the pan genome analysis of coding regions in *E. faecalis*

2) *E. faecium*

Pan genome analysis of all coding regions of *E faecium* identified a total of 7488 unique protein coding sequences found collectively in the isolates. Based on their prevalence in the genomes, the genes were identified as part of the core genome if they were present in all isolates. 2134 protein coding sequences constituted the core genome of *E. faecium*. Genes not present in all isolates but shared by most (prevalence of 15-95%) designated as Shell genes with1344 protein coding sequences belonging to this category in the generated *E. faecalis* pan genome. Even less prevalent were the Cloud genes present in only 0-15% of the genomes of *E. faecium*. These genes constitute the mobilome of *E faecium* and allow for great genomic plasticity. (Figure 8)

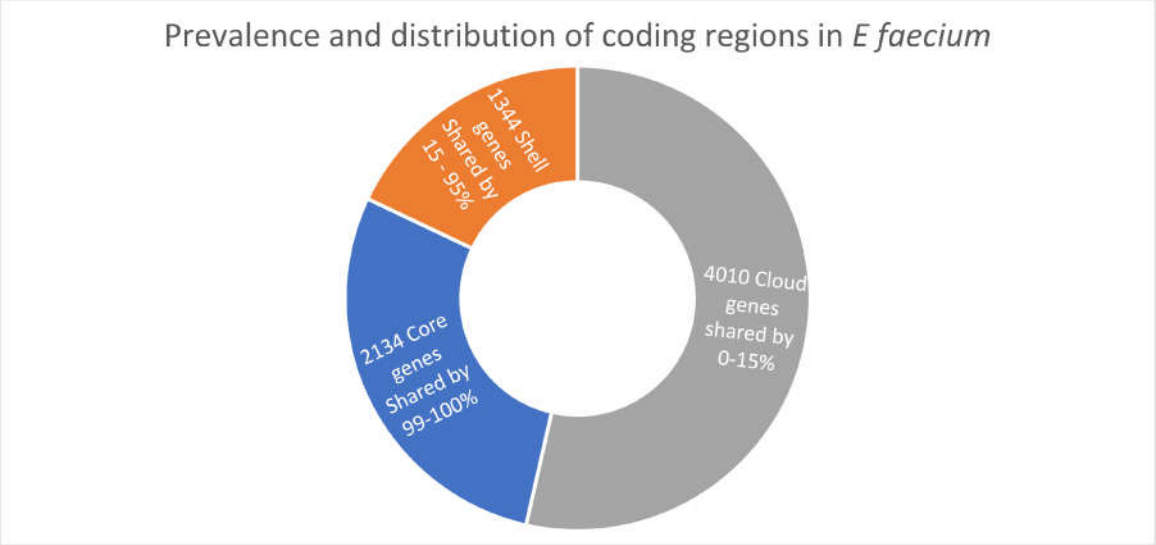


Figure 8. Circular diagram showing the prevalence and distribution of coding regions in *E.s faecium* isolates of our study.

Figure 9 illustrates the relatedness of the isolates based on shared protein coding sequences. Each blue block represents a gene, and isolates are grouped based on the number of shared/unique genes as a phylogenetic tree on the left side of the matrix.

Isolates EF8, EF9 and EF11 shared two important fragments encoding lipopolysaccharide biosynthesis protein (WzxC) and a transposase from transposon Tn916 flanking hypothetical proteins. Isolate EF6 and EF15 on the other hand, uniquely carried a putative transposon Tn552 DNA-invertase, antiporters for cadmium, cobalt, and zinc, as well as an HTH-type transcriptional repressor (CzrA). EF15 additionally encoded the vancomycin resistance cassette that is absent from the EF6 genome, as well as a 20U general stress protein. EF21 however, contained a plethora of unique genes that placed it apart from the other isolates, two of which were a membrane protein insertase (YidC), and a quinolone resistance protein (NorB).

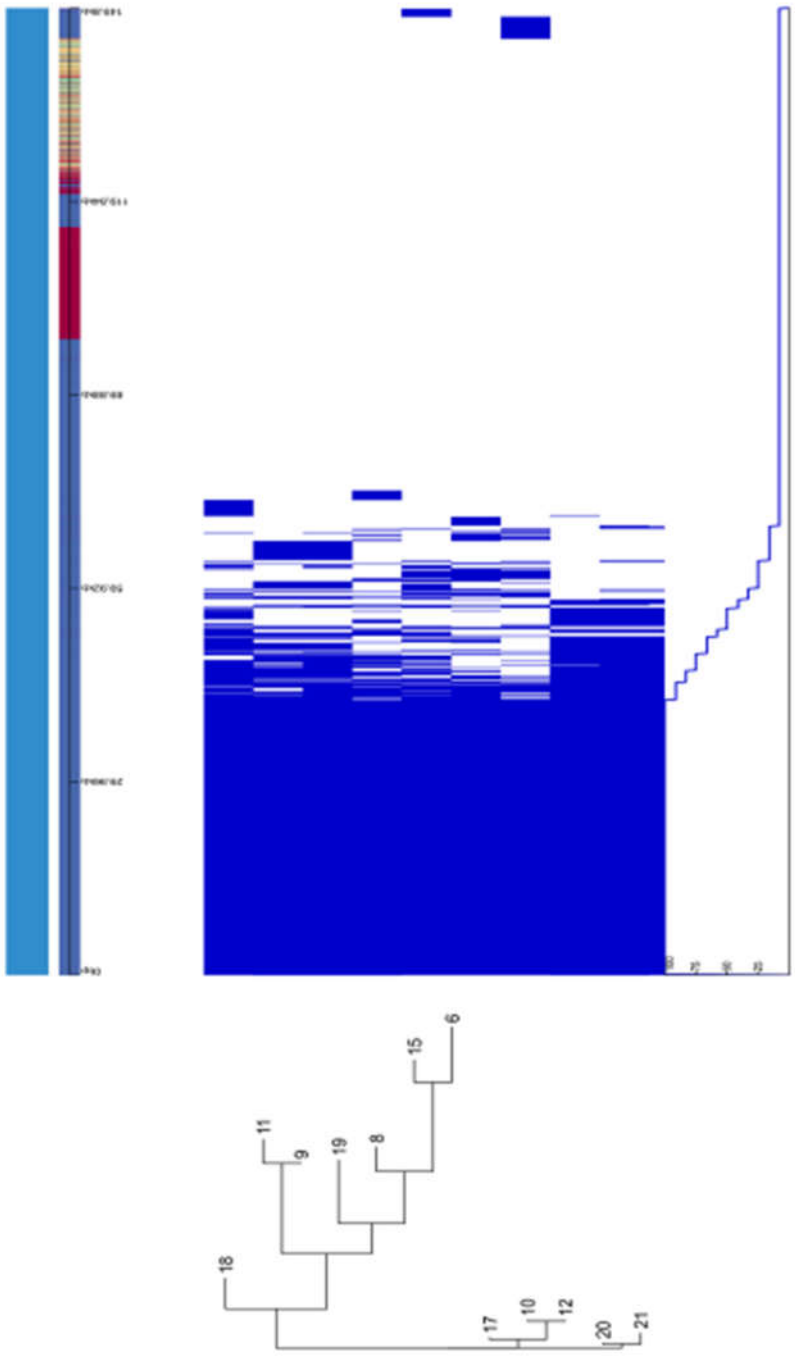


Figure 9. Presence/Absence matrix with its associated maximum likelihood phylogenetic tree based on the pan genome analysis of coding regions in *E. faecalis.*

B. Pan Genomes of Non-Coding Intergenic (IGR) Regions

1) *E. faecalis*

Figure 10 illustrates the relatedness of the isolates based on shared intergenic (IGR) sequences flanking the protein coding sequences. Each blue block represents an intergenic fragment, and isolates are grouped based on the number of shared/unique genes as a phylogenetic tree on the left side of the matrix.

The intergenic regions of the *E. faecalis* pan genome are extremely diverse. Although the coding regions outnumber the non-coding fragments and dictate whole genome based phylogenetic analyses, the non-coding IGR regions are under different levels of evolutionary stress. Accordingly, the intergenic regions harbor SNPs that are unique to each isolate. The core genes identified above in the pan genome of coding regions also have unchanged intergenic regions. Most of the observed diversity in the intergenic regions is localized around coding regions of the enterococcal mobilome.

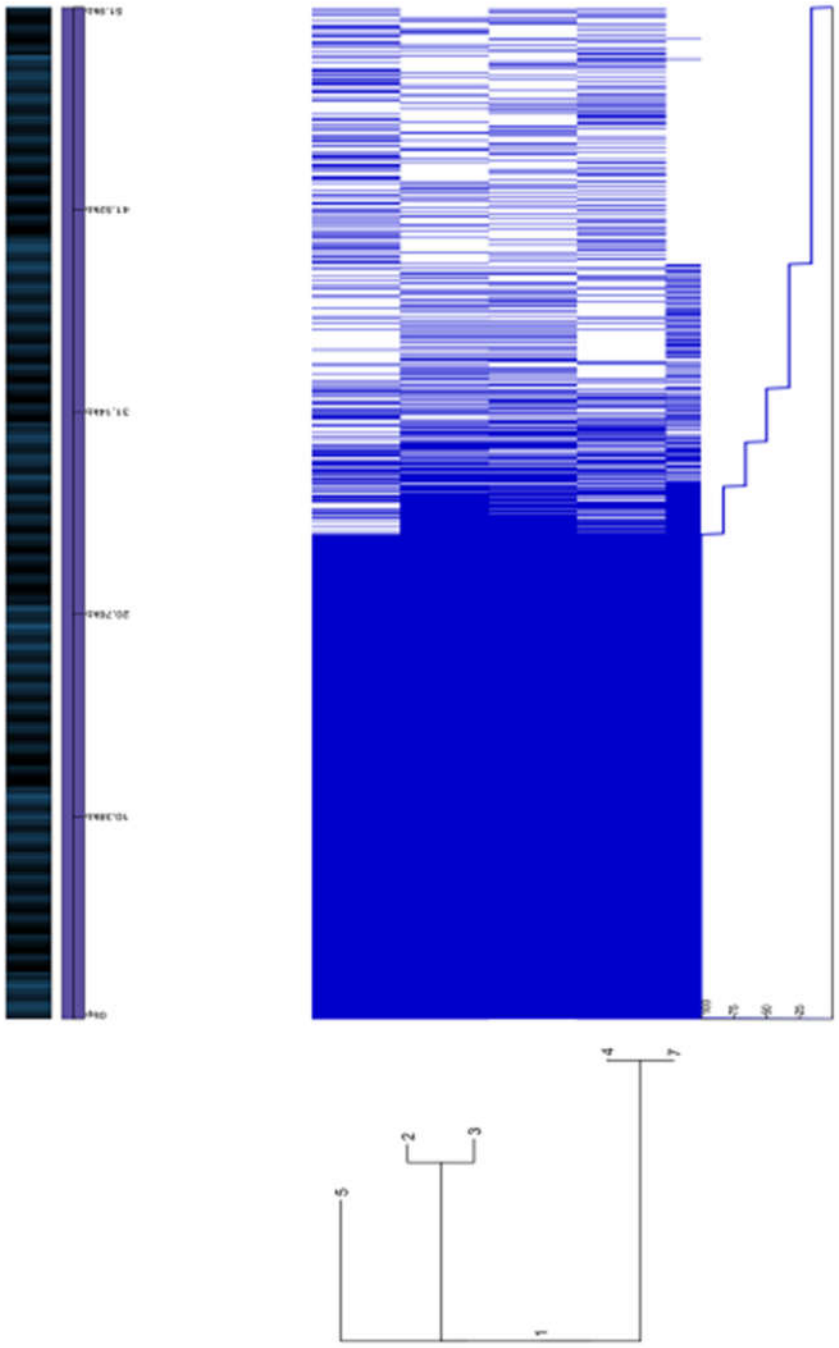


Figure 10. Presence/Absence matrix with its associated maximum likelihood phylogenetic tree based on the pan genome analysis of Non-Coding (IGR) regions in *Enterococcus faecalis*.

2) *Enterococcus faecium*

Figure 11 illustrates the relatedness of the isolates based on shared intergenic (IGR) sequences flanking the protein coding sequences. Each blue block represents an intergenic fragment, and isolates are grouped based on the number of shared/unique genes as a phylogenetic tree on the left side of the matrix.

Similar to *E. faecalis,* there was pronounced heterogeneity in the pan genome of intergenic noncoding regions of *E. faecium.* Isolates clustered similarly as that generated when using coding sequences, yet show unique fragments flanking their protein coding genes. EF6 showed the highest diversity in terms of alternative IGRs between genes, having the least number of common IGRs with the other isolates. Moreover, EF21 and EF6 clustered separately from the others due to a low number of shared IGRs. Compared to the other isolates of cluster B, t EF21 contained in total more genes that have their own unique IGRs not found in the other isolates.

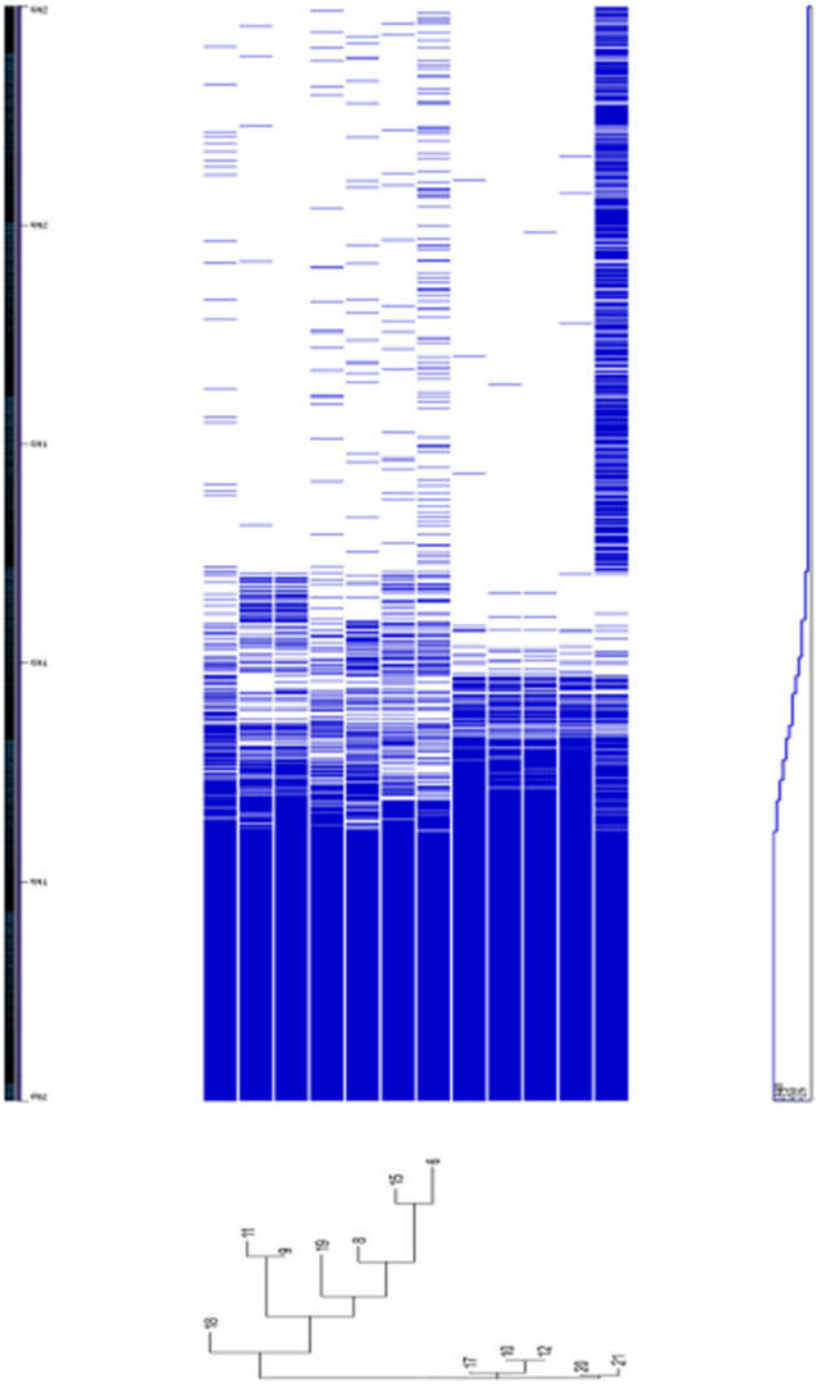


Figure 11. Presence/Absence matrix with its associated maximum likelihood phylogenetic tree based on the pan genome analysis of Non-Coding (IGR) regions in *Enterococcus faecium*

5. Recombination analysis

Recombination analysis in the core genomes of the isolates was illustrated by a presence/absence matrix coupled to its maximum likelihood phylogenetic tree. The blue blocks represent recombination regions unique to a single isolate, and the red blocks represent recombination loci shared by multiple isolates. The blocks span the length of the recombination hotspot according to their genomic position on the X axis. Metadata of the recombination analyses (Total SNPs, number of SNPs inside recombinations, number of SNPs outside recombinations, recombination blocks, bases in recombinations, r/m values, rho/theta, genome length, and bases in clonal frame) were annotated as colored blocks along the level of each isolate. Similar to the pan genome, the results of the recombination analysis were too numerous to be discussed each on its own, and as a result we again opted to select the most prevalent recombination loci and identified the relevant proteins.

A. E. faecalis (Figure 12)

Three large recombination hotspots were examined. The largest block (A) representing a 100KB locus spanning bases 1,650,000 – 1,750,000 of the genomes, was highly polymorphic for isolates EF2 and EF3. This block contained coding regions of clinical significance, including but not limited to Penicillin binding proteins, uracil permease, formate-tetrahydrofolate ligase, proline dipeptidase, phosphate ABC transporter (permease), cell division ATP-binding protein FtsE, preprotin translocase SecA, phosphoribosylaminoimidazolecarboxamide formyltransferase (IMP cyclohydrolase), phosphoribosylformylglycinamidine synthase II, and a pheromone binding protein.

A second highly recombinant locus (B) around position 1,952,450 of the genomes contained site specific recombinase (phage integrase family), and an altered ABC transporter (permease)

The third favored locus of recombination (C) contained inositol monophosphatase, and many hypothetical proteins, some of which were truncated or were uncharacterized.

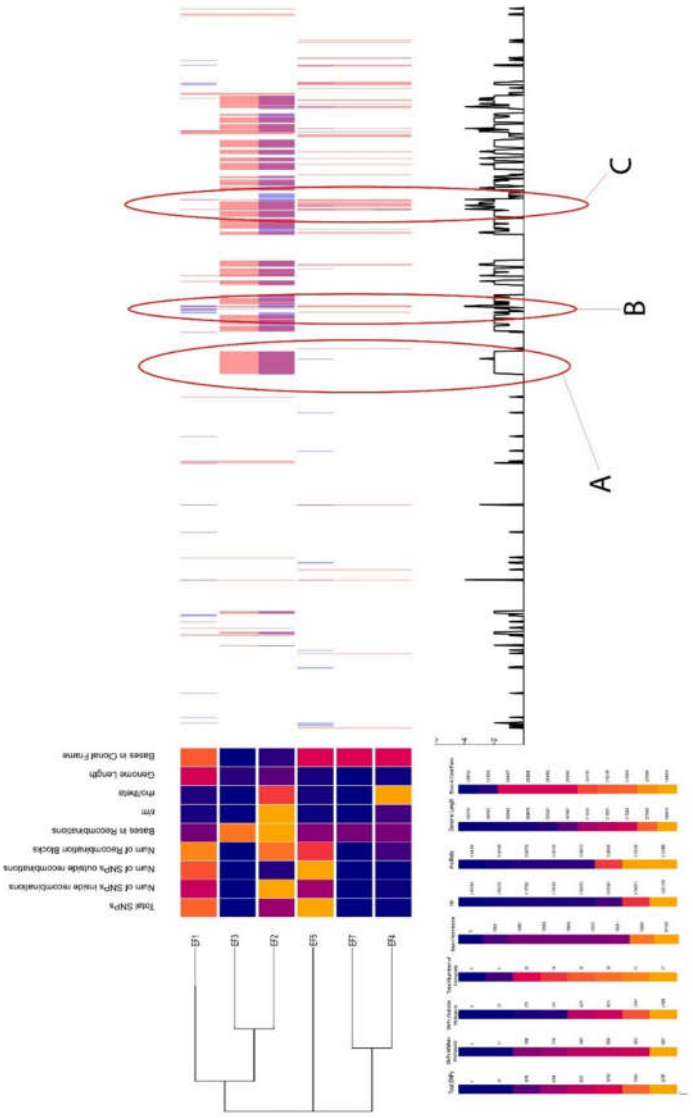


Figure 12. Presence/Absence matrix with its associated maximum likelihood phylogenetic tree based on the core genome recombinations in *Enterococcus faecalis*. Red bars show recombination loci shared between isolates. Blue bars show recombination loci unique to an isolate.

B. *E. faecium* (Figure 13)

The recombinations of *E. faecium* isolates were organized into short loci. Five representative hotspots with the highest recombination rates were analyzed. The hotspot observed around 831,560 of the genomes (A) contained many fragments encoding both structural and functional subunits of bacteriophages, an S14 family endopeptidase ClpP, along with few hypothetical proteins that were yet to be characterized. The hotspot observed around genomic locus 1,250,000 (B) harbored regions encoding transketolase, glyceraldehyde-3-phosphate dehydrogenase, and hypothetical proteins, some of which were linked to recombinogenic phenotypes. A large block of 120kb (C) spanning the 1,570,000 – 1,690,000 region contained fragments encoding DNA helicase, methyltransferases, an excision endonuclease subunit UvrA, and a microcin C7 resistance protein. The region around 1,910,000 (D) was also highly polymorphic containing genes linked to survival in harsh environments, including but not limited to penicillin-binding protein 1b, IS66 transposase, a recombination protein RecO, and a multidrug ABC superfamily ATB binding cassette transporter. The fifth hotspot (E) at around locus 2,620,000 included DNA modification genes such as a highly polymorphic DNA mismatch repair protein MutS, and a DNA-directed RNA polymerase subunit beta (rpoB)

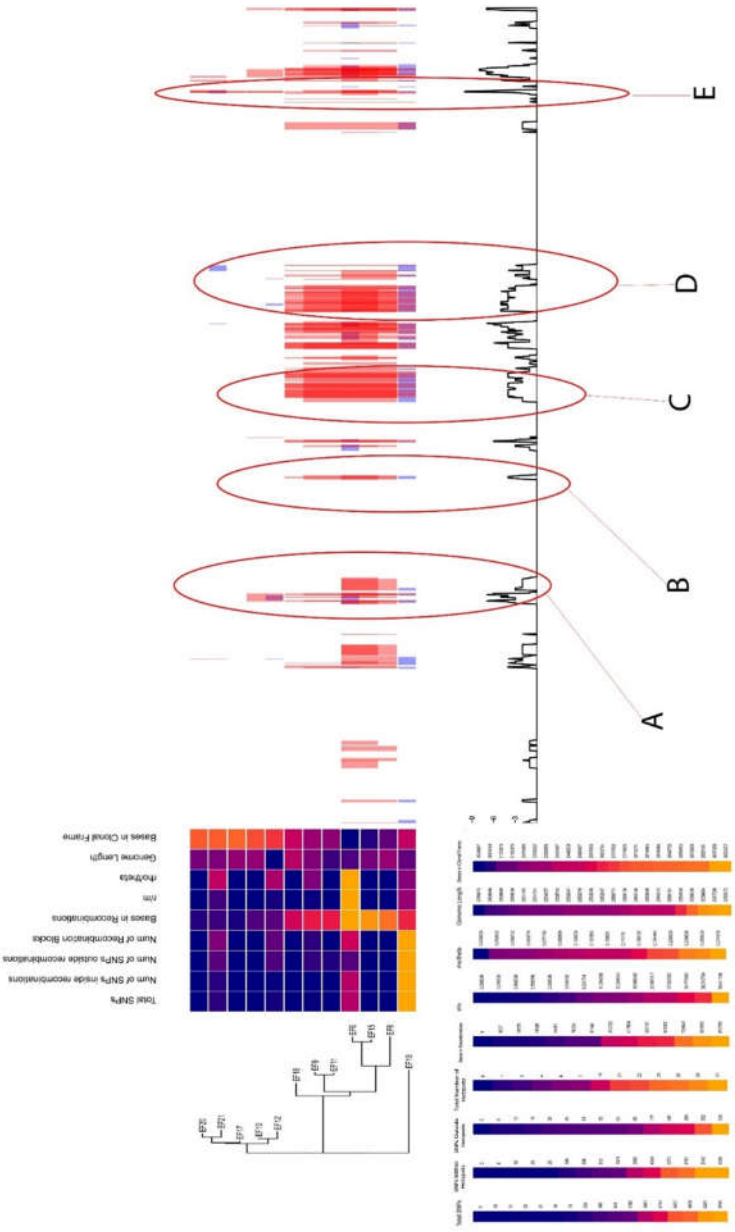


Figure 13. Presence/Absence matrix with its associated maximum likelihood phylogenetic tree based on the core genome recombinations in *Enterococcus faecium*. Red bars show recombination loci shared between isolates; Blue bars show recombination loci unique to an isolate.

Chapter IV

Discussion

Enterococci have thrived under the harsh circumstances of nature by rapidly adapting to withstand elevated levels of selective pressure. Enterococci boast an extremely hardened phenotype and continue to introduce polymorphisms that increase the genetic complexity and pathogenicity.

In this study the studied enterococci showed very diverse multilocus sequence types, being mainly typed as ST 80 8/21 (38%) and ST 203 5/21 (24%), in addition to detecting ST 28, ST 40, ST 55, ST 117, ST 376, and ST 480. Some of these isolates such as the ST80 was previously associated with epidemics linked to highly resistant enterococci (Pinholt, 2016). As a result of the observed hypermutability of enterococcal genomes, some showed altered housekeeping genes giving rise to new ST types, or even at times the re-emergence of ancient ST types through SNPs localized in short defined fragments of recombination hotspots (van Hal et al, 2016).

Enterococci have embraced genomic plasticity to shuffle their genomes leading to the generation of more resistant phenotypes. They achieve this through the generation of single nucleotide variants, but more commonly they share key phenotypic traits *en block* via horizontal gene transfer using transposons and a few plasmids (Palmer L, 2010). The mobilization of resistance cassettes, in particular the vancomycin resistance operon, was observed between the almost identical isolates EF 6 (vancomycin sensitive) and EF 15 (vancomycin resistant). This further emphasizes on the need for more accurate and in-depth analysis methods to overcome the pitfalls of classical phenotypic testing. By treating with vancomycin,

we are further promoting the transfer of vancomycin resistance to previously susceptible enterococci that are highly similar to circulating VRE clones.

Standard laboratory procedures to characterize enterococcal infections are limited to a few phenotypic antimicrobial susceptibility tests and source of isolation of the bacterium (Table 3). In contrast, whole genome sequence-based resistance profiling approaches give more informative results by aligning genes of interest to international references and to each other. Tables 4 and 5 show the genetic basis of each of the resistance phenotypes observed in Table 3. Isolates are characterized based on the presence of specific resistance determinants that even though code for the same phenotype, but are different in their levels of tolerance, transmissibility, and even geographic prevalence. These results are important for outbreak investigations and improved patient treatment protocols.

Recombination plays a key role in increasing the genomic heterogeneity of enterococci. This gives them the needed advantage to survive under the selective pressures of the hospital environment such as disinfectants, bactericidal agents, and antimicrobials. Our isolates showed highly active recombination hotspots (Figure 12, Figure 13) with genes that when recombined give altered phenotypes. One example of this are the altered penicillin binding proteins that were found in highly polymorphic loci of recombination. By altering their binding proteins (Figure 12 – Block A) , genes encoding cell wall components, and pumps (Figure 13 – Block D) that govern the flux of antimicrobials, enterococci decrease their susceptibility to the commonly used antibiotics by decreased binding affinity, decreased influx, or highly active export systems.

As naturally occurring gut commensals, enterococci also colonized the gut of insects that feed on plants, soil, and especially dead tree barks. According to Sepike et al, rotting plant matter contains antibiotic producing Streptomycetes. Ingestion of the plant matter along with Streptomycetes puts these antibiotic producers in close contact with the enterococci of the gut. Hence, the increased selective pressure in the gut of the insect forces microevolutionary events triggering antibiotic resistance phenotypes in enterococci. (Sepike et al, 2012). Our results indicate that the sudden emergence of vancomycin resistance in Lebanon also followed a similar evolutionary path. Isolates with different sources of isolation (Figure 3) and different ST types (Table 5) shared a common vancomycin resistance phenotype. This unrestricted propagation of the vancomycin resistance cassette into unrelated clones such as EF 1 (ST 28), EF 2 (ST 55), EF 3 (ST 40), EF 4 (ST 480), EF 5 (ST 376), EF 6 (ST 80), EF 10 (ST 203), EF 13 (ST 224), and EF 18 (ST 117) is highly alarming as it devalues the classical multilocus typing scheme.

To negate the idea of de novo generation of antibacterial resistance genes in enterococci that are under high selective pressure, Benveniste et al showed that the aminoglycoside inactivating enzymes of enterococci are highly similar to those found in actinomycetes, suggesting that these fragments were transferred at some point during evolution into enterococci (Benveniste et al, 1973). Even though our isolates have highly recombining genomes, it is unlikely that they generated vancomycin resistance solely through genetic alterations due to their identical vancomycin resistance genes (Table 5). D'costa et al showed high relatedness between the resistance determinants found in multidrug resistant enterococci and other bacteria resistant to the same antimicrobials. It is imperative that we also study the resistance determinants of other gram-positive bacteria in the region to account

for the emergence of novel resistant clones in the future. Isolate EF 21 uniquely showed resistance determinants to beta lactams, fusidic acid, and additional aminoglycoside resistance genes *ant(4')-la* and *aadD2*. More commonly found in staphylococci (Chang, 2003) such resistance genes are transposed onto enterococci only when under selective pressure induced by prolonged usage of the same antimicrobials.

The observed resistance phenotypes (Table 3) were encoded by more than one gene per phenotype (Table 5), demonstrating how enterococci have gradually collected their resistance determinants in Lebanon, as opposed to the propagation of one successful epidemic clone that would be easier to track and manage in the clinical setting.

The famed *vanA* vancomycin operon is much older than its first appearance in the clinical setting. Molecular identification of the *vanA* operon homology and ancestry showed how it was derived from the great ancestor *Amycolatopsis orientalis* (D'costa et al, 2011). Vancomycin resistance was not a highly advantageous phenotype in ancient times where selective pressure on the enterococcal cells depended solely on their surrounding cells and environment. With the extensive use of antibacterial agents, and in particular vancomycin, we have seen a boom in acquired antibiotic resistance phenotypes such as the VanA and VanB phenotypes in enterococci. Depending on how we manage enterococcal infections in the future, the vancomycin resistance phenotype is also expected to shift its genetic basis for increased fitness. Additionally, future studies should explore the existence of alternate vancomycin resistance cassettes other than Van-A and Van-B in non-clinical isolates (Gelsomino 2004, Serio 2007).

Along with the ancient genetic origins of vancomycin, its associated resistance phenotype also dates back centuries. Enterococcus casseliflavus present in 45% of the gut of insects harbors a vancomycin resistance operon highly similar to that of the VanA phenotype observed in our isolates as well as in international clones (Cattoir et al, 2012). Future studies could also explore the gut metagenome of insects that harbor enterococci to comparatively analyze the vancomycin resistance cassette.

Similarly, the impact of humans on vancomycin resistance came from the misuse of Avoparcin, which is highly similar in structure to vancomycin. Avoparcin was extensively used in Europe for agricultural and farming purposes. Studies investigating the role of avoparcin in contributing to increased resistance to vancomycin in enterococci revealed the high correlation of these two variables. As the usage of Avoparcin increases in a certain geographical area, the chance for a successful VRE clone to emerge is quite plausible. Banning the usage of avoparcin in Denmark (Bager et al, 1999) and France (2008) decreased VRE incidence rates but only of the non-hospital acquired enterococci. To our knowledge, no study has elucidated the usage of avoparcin and its effects on vancomycin resistance in enterococci in the region.

Enterococci do not seem to evolve based on their geographical niche, but rather according to what surrounds them for a short period of time in their evolutionary path. Interestingly, based on a higher than average usage of vancomycin in the clinical setting, Luxembourg had an average of 10 times more vancomycin resistant enterococcal infections compared to its neighboring countries (Cattoir, 2012). In concordance with this, previously characterized enterococci of our region such as Saudi Arabia and Iran show pronounced heterogeneity in their genes encoding

vancomycin resistance. Our characterized isolates also show clonal diversity and uniqueness in their genomic makeup.

The pan genome analysis was initially performed on the whole collection as one large group but gave unsatisfying results due to the low number of core genes shared by all isolates. Hence, analysis of *Enterococcus faecalis* and *Enterococcus faecium* isolates on their own gave more comprehensive results. *E faecalis* isolates had 2263 core genes, while *E faecium* isolates had 2134 core genes in accordance with a similar study (Raven, 2016). The main differences between the pan genomes of these two species are in the regions not part of the core genomes. The shell and accessory genomes encode for a plethora of proteins not common in all isolates. A large portion of these accessory genes were either truncated or annotated as hypothetical proteins due to the lack of evidence. Both accessory pan genomes contained invertases, integrases, translocases, and DNA remodeling proteins (Figure 7 and Figure 9) that account for the high recombination rates seen in their core genomes (Figure 12 and Figure 13).

Enterococci looking to acquire genes through horizontal transfer have a wide variety of options when found in a complex microbial environment. These highly diverse bacterial superpopulations are able to supply genetic fragments to the rapidly evolving enterococci and increase its diversity both in nature and in the clinical setting (Murray, 1998).

With increasing flexibility in their genetic transpositions, logic follows the progression of vancomycin resistance from enterococci to other bacteria nearing the same stage of saturated treatment options. Most importantly the emergence of vancomycin resistant staphylococcus aureus (VRSA) is a huge concern to the

healthcare professionals. The high plasticity of the enterococcal genomes threatens to transmit the vancomycin resistance determinants to its close relative, *Staphylococcus aureus*. Nevertheless, there have been no studies that show direct transmissibility of the vancomycin resistance operon, as well as the vancomycin resistance phenotype, from enterococci to staphylococci despite high genetic similarity. (Chang, 2003; Kos, 2012; Weigel, 2003)

With the advent of vancomycin resistance in Lebanon, it is crucial to study the genomic intricacies of these rapidly evolving pathogens. Using whole genome sequence based bioinformatic analyses, we were able to infer phylogeny based on core genome SNPs found in genes that are ubiquitous in all tested isolates. Phylogeny based on assemblies gave poor results that could not give results more informative than species identification and had low discriminatory power. We were able to achieve high discriminatory power by mapping the whole genome sequence reads directly onto a reference genome and selecting for high quality single nucleotide variants. These alignments were useful in reconstructing phylogenetic trees and assessing recombination hotspots in the genomes.

The recombination analysis also revealed the highly active polymorphic sites that govern some of the genomic plasticity of enterococci. Pan genome analysis also revealed the flexibility of isolates in sharing genetic fragments, be it small or large sized.

A collection of yet uncharacterized hypothetical proteins was found both in the pan genome of coding regions, as well as in certain recombination loci, likely pointing towards a novel function in survival or antibacterial resistance.

To our knowledge, this is the first study to perform an end to end whole genome sequence based bioinformatic analysis of the genomes of circulating VRE clones in the region.

References

Alanis, A. J. (2005). Resistance to antibiotics: are we in the post-antibiotic era?. Archives of medical research, 36(6), 697-705.

Andrewes, F. W., & Horder, T. (1906). A study of the streptococci pathogenic for man. The Lancet, 168(4334), 775-783.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.

Arias, C. A., & Murray, B. E. (2012). The rise of the Enterococcus: beyond vancomycin resistance. Nature Reviews Microbiology, 10(4), 266.

Arthur, M. I. C. H. E. L., Molinas, C., Depardieu, F., & Courvalin, P. (1993). Characterization of Tn1546, a Tn3-related transposon conferring glycopeptide resistance by synthesis of depsipeptide peptidoglycan precursors in Enterococcus faecium BM4147. Journal of bacteriology, 175(1), 117-127.

Ashkenazy, H., Penn, O., Doron-Faigenboim, A., Cohen, O., Cannarozzi, G., Zomer, O., & Pupko, T. (2012). FastML: a web server for probabilistic reconstruction of ancestral sequences. Nucleic acids research, 40(W1), W580-W584.

Bager, F., Aarestrup, F. M., Madsen, M., & Wegener, H. C. (1999). Glycopeptide resistance in Enterococcus faecium from broilers and pigs following discontinued use of avoparcin. Microbial Drug Resistance, 5(1), 53-56.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., ... & Pyshkin, A. V. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of computational biology, 19(5), 455-477.

Ben Said, L., Klibi, N., Dziri, R., Borgo, F., Boudabous, A., Ben Slama, K., & Torres, C. (2016). Prevalence, antimicrobial resistance and genetic lineages of Enterococcus spp. from vegetable food, soil and irrigation water in farm environments in Tunisia. Journal of the Science of Food and Agriculture, 96(5), 1627-1633.

Bender, J. K., Kalmbach, A., Fleige, C., Klare, I., Fuchs, S., & Werner, G. (2016). Population structure and acquisition of the vanB resistance determinant in German clinical isolates of Enterococcus faecium ST192. Scientific reports, 6, 21847.

Benveniste, R., & Davies, J. (1973). Aminoglycoside antibiotic-inactivating enzymes in actinomycetes similar to those present in clinical isolates of antibiotic-resistant bacteria. Proceedings of the National Academy of Sciences, 70(8), 2276-2280.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), 2114-2120.

Bonacina, J., Suárez, N., Hormigo, R., Fadda, S., Lechner, M., & Saavedra, L. (2016). A genomic view of food-related and probiotic Enterococcus strains. Dna Research, 24(1), 11-24.

Bourgeois-Nicolaos, N., Nguyen, T. T., Defrance, G., Massias, L., Alavoine, L., Lefort, A., ... & Andremont, A. (2014). The emergence of linezolid resistance among Enterococci in intestinal microbiota of treated patients is unrelated to individual pharmacokinetic characteristics. Antimicrobial agents and chemotherapy, 58(5), 2681-2687.

Boyce, J. M., Mermel, L. A., Zervos, M. J., Rice, L. B., Potter-Bynoe, G., Giorgio, C., & Medeiros, A. A. (1995). Controlling vancomycin-resistant enterococci. Infection Control & Hospital Epidemiology, 16(11), 634-637.

Brodrick, H. J., Raven, K. E., Harrison, E. M., Blane, B., Reuter, S., Török, M. E., ... & Peacock, S. J. (2016). Whole-genome sequencing reveals transmission of vancomycin-resistant Enterococcus faecium in a healthcare network. Genome medicine, 8(1), 4.

Buultjens, A. H., Lam, M. M., Ballard, S., Monk, I. R., Mahony, A. A., Grabsch, E. A., ... & Seemann, T. (2017). Evolutionary origins of the emergent ST796 clone of vancomycin resistant Enterococcus faecium. PeerJ, 5, e2916.

Caballero, S., Kim, S., Carter, R. A., Leiner, I. M., Sušac, B., Miller, L., ... & Pamer, E. G. (2017). Cooperating commensals restore colonization resistance to vancomycin-resistant Enterococcus faecium. Cell host & microbe, 21(5), 592-602.

Cattoir, V., & Leclercq, R. (2012). Twenty-five years of shared life with vancomycin-resistant enterococci: is it time to divorce? Journal of Antimicrobial Chemotherapy, 68(4), 731-742.

Chang, S., Sievert, D. M., Hageman, J. C., Boulton, M. L., Tenover, F. C., Downes, F. P., ... & Cardo, D. (2003). Infection with vancomycin-resistant Staphylococcus aureus containing the vanA resistance gene. New England Journal of Medicine, 348(14), 1342-1347.

Courvalin, P. (2006). Vancomycin resistance in gram-positive cocci. Clinical Infectious Diseases, 42(Supplement_1), S25-S34.

Croucher, N. J., Page, A. J., Connor, T. R., Delaney, A. J., Keane, J. A., Bentley, S. D., ... & Harris, S. R. (2014). Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic acids research, 43(3), e15-e15.

Crouzet, L., Rigottier-Gois, L., & Serror, P. (2015). Potential use of probiotic and commensal bacteria as non-antibiotic strategies against vancomycin-resistant enterococci. FEMS microbiology letters, 362(8).

D'Costa, V. M., King, C. E., Kalan, L., Morar, M., Sung, W. W., Schwarz, C., ... & Golding, G. B. (2011). Antibiotic resistance is ancient. Nature, 477(7365), 457.

Davies, J. (1990). What are antibiotics? Archaic functions for modern activities. Molecular microbiology, 4(8), 1227-1232.

Davies, J., & Davies, D. (2010). Origins and evolution of antibiotic resistance. Microbiol. Mol. Biol. Rev., 74(3), 417-433.

D'costa, V. M., McGrann, K. M., Hughes, D. W., & Wright, G. D. (2006). Sampling the antibiotic resistome. Science, 311(5759), 374-377.

Deshpande, L. M., Ashcraft, D. S., Kahn, H. P., Pankey, G., Jones, R. N., Farrell, D. J., & Mendes, R. E. (2015). Detection of a new cfr-like gene, cfr (B), in Enterococcus faecium isolates recovered from human specimens in the United States as part of the SENTRY Antimicrobial Surveillance Program. Antimicrobial agents and chemotherapy, 59(10), 6256-6261.

Donskey, C. J., Chowdhry, T. K., Hecker, M. T., Hoyen, C. K., Hanrahan, J. A., Hujer, A. M., ... & Rice, L. B. (2000). Effect of antibiotic therapy on the density of vancomycin-resistant enterococci in the stool of colonized patients. New England Journal of Medicine, 343(26), 1925-1932.

Facklam, R. R., Maria da Gloria, S. C., & Teixeira, L. M. (2002). History, taxonomy, biochemical characteristics, and antibiotic susceptibility testing of enterococci. In The Enterococci (pp. 1-54). American Society of Microbiology.

Fatholahzadeh, B., HASHEMI, F. B., EMANEINI, M., ALIGHOLI, M., NAKHJAVANI, F. A., & KAZEMI, B. (2006). Detection of vancomycin resistant enterococci (VRE) isolated from urinary tract infections (UTI) in Tehran, Iran. DARU Journal of Pharmaceutical Sciences, 14(3), 141-145.

Franz, C. M., Stiles, M. E., Schleifer, K. H., & Holzapfel, W. H. (2003). Enterococci in foods—a conundrum for food safety. International journal of food microbiology, 88(2-3), 105-122.

Gelsomino, R., Vancanneyt, M., Vandekerckhove, T. M., & Swings, J. (2004). Development of a 16S rRNA primer for the detection of Brevibacterium spp. Letters in applied microbiology, 38(6), 532-535.

Giraffa, G. (2002). Enterococci from foods. FEMS microbiology reviews, 26(2), 163-171.

Goh, S. H., Facklam, R. R., Chang, M., Hill, J. E., Tyrrell, G. J., Burns, E. C., ... & Hemmingsen, S. M. (2000). Identification of Enterococcus Species and Phenotypically Similar Lactococcus andVagococcus Species by Reverse Checkerboard Hybridization to Chaperonin 60 Gene Sequences. Journal of clinical microbiology, 38(11), 3953-3959.

Gouliouris, T., Raven, K. E., Ludden, C., Blane, B., Corander, J., Horner, C. S., ... & Holmes, M. A. (2018). Genomic surveillance of Enterococcus faecium reveals limited sharing of strains and resistance genes between livestock and humans in the United Kingdom. MBio, 9(6), e01780-18.

Guzman Prieto, A. M., van Schaik, W., Rogers, M. R., Coque, T. M., Baquero, F., Corander, J., & Willems, R. J. (2016). Global emergence and dissemination of enterococci as nosocomial pathogens: attack of the clones? Frontiers in microbiology, 7, 788.

Hadfield, J., Croucher, N. J., Goater, R. J., Abudahab, K., Aanensen, D. M., & Harris, S. R. (2017). Phandango: an interactive viewer for bacterial population genomics. Bioinformatics, 34(2), 292-293.

Hegstad, K., Mikalsen, T., Coque, T. M., Werner, G., & Sundsfjord, A. (2010). Mobile genetic elements and their contribution to the emergence of antimicrobial resistant Enterococcus faecalis and Enterococcus faecium. Clinical microbiology and infection, 16(6), 541-554.

Homan, W. L., Tribe, D., Poznanski, S., Li, M., Hogg, G., Spalburg, E., ... & Willems, R. J. (2002). Multilocus sequence typing scheme for Enterococcus faecium. Journal of clinical microbiology, 40(6), 1963-1971.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution, 30(4), 772-780.

Kempf, I., Hellard, G., Perrin-Guyomard, A., Gicquel-Bruneau, M., Sanders, P., & Leclercq, R. (2008). Prevalence of high-level vancomycin-resistant enterococci in French broilers and pigs. International journal of antimicrobial agents, 32(5), 463-464.

Khan, M. A., Van Der Wal, M., Farrell, D. J., Cossins, L., Van Belkum, A., Alaidan, A., & Hays, J. P. (2008). Analysis of VanA vancomycin-resistant Enterococcus faecium isolates from Saudi Arabian hospitals reveals the presence of clonal cluster 17 and two new Tn 1546 lineage types. Journal of Antimicrobial Chemotherapy, 62(2), 279-283.

Kopit, L. M., Kim, E. B., Siezen, R. J., Harris, L. J., & Marco, M. L. (2014). Safety of the surrogate microorganism Enterococcus faecium NRRL B-2354 for use in thermal process validation. Appl. Environ. Microbiol., 80(6), 1899-1909.

Kos, V. N., Desjardins, C. A., Griggs, A., Cerqueira, G., Van Tonder, A., Holden, M. T., ... & Wortman, J. (2012). Comparative genomics of vancomycin-resistant Staphylococcus aureus strains and their positions within the clade most commonly associated with Methicillin-resistant S. aureus hospital-acquired infection in the United States. MBio, 3(3), e00112-12.

Kuroda, M., Sekizuka, T., Matsui, H., Suzuki, K., Seki, H., Saito, M., & Hideaki, H. A. N. A. K. I. (2018). Complete Genome Sequence and Characterization of Linezolid-resistant Enterococcus faecalis Clinical Isolate KUB3006 carrying a cfr (B)-

Transposon on its chromosome and optrA-Plasmid. Frontiers in microbiology, 9, 2576.

Launay, A., Ballard, S. A., Johnson, P. D., Grayson, M. L., & Lambert, T. (2006). Transfer of vancomycin resistance transposon Tn1549 from Clostridium symbiosum to Enterococcus spp. in the gut of gnotobiotic mice. Antimicrobial agents and chemotherapy, 50(3), 1054-1062.

Leavis, H. L., Bonten, M. J., & Willems, R. J. (2006). Identification of high-risk enterococcal clonal complexes: global dispersion and antibiotic resistance. Current opinion in microbiology, 9(5), 454-460.

Lebreton, F., van Schaik, W., McGuire, A. M., Godfrey, P., Griggs, A., Mazumdar, V., ... & Zeng, Q. (2013). Emergence of epidemic multidrug-resistant Enterococcus faecium from animal and commensal strains. MBio, 4(4), e00534-13.

Leong, K. W., Cooley, L. A., Anderson, T. L., Gautam, S. S., McEwan, B., Wells, A., ... & O'Toole, R. F. (2018). Emergence of vancomycin-resistant enterococcus faecium at an Australian hospital: a whole genome sequencing analysis. Scientific reports, 8(1), 6274.

Letunic, I., & Bork, P. (2006). Interactive Tree of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics, 23(1), 127-128.

López, M., Sáenz, Y., Rojo-Bezares, B., Martínez, S., del Campo, R., Ruiz-Larrea, F., ... & Torres, C. (2009). Detection of vanA and vanB2-containing enterococci from food samples in Spain, including Enterococcus faecium strains of CC17 and the new singleton ST425. International journal of food microbiology, 133(1-2), 172-178.

Löytynoja, A. (2014). Phylogeny-aware alignment with PRANK. In Multiple sequence alignment methods (pp. 155-170). Humana Press, Totowa, NJ.

MacCallum, W. G., & Hastings, T. W. (1899). A case of acute endocarditis caused by Micrococcus zymogenes (nov. spec.), with a description of the microorganism. Journal of Experimental Medicine, 4(5-6), 521-534.

Mason, K. L., Stepien, T. A., Blum, J. E., Holt, J. F., Labbe, N. H., Rush, J. S., ... & Handelsman, J. (2011). From commensal to pathogen: translocation of Enterococcus faecalis from the midgut to the hemocoel of Manduca sexta. MBio, 2(3), e00065-11.

Méndez-Álvarez, S., Pérez-Hernández, X., & Claverie-Martín, F. (2000). Glycopeptide resistance in enterococci. International Microbiology, 3(2), 71-80.

Monstein, H. J., Tiveljung, A., Kraft, C. H., Borch, K., & Jonasson, J. (2000). Profiling of bacterial flora in gastric biopsies from patients with Helicobacter pylori-associated gastritis and histologically normal control individuals by temperature gradient gel electrophoresis and 16S rDNA sequence analysis. Journal of medical microbiology, 49(9), 817-822.

Murray, B. E. (1998). Diversity among multidrug-resistant enterococci. Emerging infectious diseases, 4(1), 37.

Naimi, A., Beck, G., Monique, M., Lefèbvre, G., & Branlanti, C. (1999). Determination of the nucleotide sequence of the 23S ribosomal RNA and flanking spacers of an Enterococcus faecium strain, reveals insertion-deletion events in the ribosomal spacer 1 of enterococci. Systematic and applied microbiology, 22(1), 9-21.

Ozawa, Y., Courvalin, P., & Galimand, M. (2000). Identification of enterococci at the species level by sequencing of the genes for D-alanine: D-alanine ligases. Systematic and applied microbiology, 23(2), 230-237.

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T., ... & Parkhill, J. (2015). Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics, 31(22), 3691-3693.

Palmer, K. L., Godfrey, P., Griggs, A., Kos, V. N., Zucker, J., Desjardins, C., ... & Feldgarden, M. (2012). Comparative genomics of enterococci: variation in Enterococcus faecalis, clade structure in E. faecium, and defining characteristics of E. gallinarum and E. casseliflavus. MBio, 3(1), e00318-11.

Palmer, K. L., Kos, V. N., & Gilmore, M. S. (2010). Horizontal gene transfer and the genomics of enterococcal antibiotic resistance. Current opinion in microbiology, 13(5), 632-639.

Patel, R., Piper, K. E., Rouse, M. S., Steckelberg, J. M., Uhl, J. R., Kohner, P., ... & Kline, B. C. (1998). Determination of 16S rRNA sequences of enterococci and application to species identification of nonmotile Enterococcus gallinarum isolates. Journal of Clinical Microbiology, 36(11), 3399-3407.

Pinholt, M., Gumpert, H., Bayliss, S., Nielsen, J. B., Vorobieva, V., Pedersen, M., ... & Westh, H. (2016). Genomic analysis of 495 vancomycin-resistant Enterococcus faecium reveals broad dissemination of a vanA plasmid in more than 19 clones from Copenhagen, Denmark. Journal of Antimicrobial Chemotherapy, 72(1), 40-47.

Pinholt, M., Larner-Svensson, H., Littauer, P., Moser, C. E., Pedersen, M., Lemming, L. E., ... & Dzajic, E. (2015). Multiple hospital outbreaks of vanA Enterococcus faecium in Denmark, 2012–13, investigated by WGS, MLST and PFGE. Journal of Antimicrobial Chemotherapy, 70(9), 2474-2482.

Poyart, C., Quesnes, G., & Trieu-Cuot, P. (2000). Sequencing the gene encoding manganese-dependent superoxide dismutase for rapid species identification of enterococci. Journal of Clinical Microbiology, 38(1), 415-418.

Quednau, M., Ahrne, S., Petersson, A. C., & Molin, G. (1998). Antibiotic-resistant strains of Enterococcus isolated from Swedish and Danish retailed chicken and pork. Journal of Applied Microbiology, 84(6), 1163-1170.

Raven, K. E., Reuter, S., Reynolds, R., Brodrick, H. J., Russell, J. E., Török, M. E., ... & Peacock, S. J. (2016). A decade of genomic history for healthcare-associated Enterococcus faecium in the United Kingdom and Ireland. Genome research, 26(10), 1388-1396.

Reuter, S., Ellington, M. J., Cartwright, E. J., Köser, C. U., Török, M. E., Gouliouris, T., ... & Parkhill, J. (2013). Rapid bacterial whole-genome sequencing to enhance diagnostic and public health microbiology. JAMA internal medicine, 173(15), 1397-1404.

Ribeiro, T., Oliveira, M., Fraqueza, M. J., Lauková, A., Elias, M., Tenreiro, R., ... & Semedo-Lemsaddek, T. (2011). Antibiotic resistance and virulence factors among enterococci isolated from Chouriço, a traditional Portuguese dry fermented sausage. Journal of food protection, 74(3), 465-469.

Santona, A., Taviani, E., Deligios, M., Al-Qahtani, A. A., Al-Ahdal, M. N., Rubino, S., & Paglietti, B. (2016). Vancomycin-resistant Enterococcus faecium high-resolution typing by core genome multilocus sequence typing. The Journal of Infection in Developing Countries, 10(10), 1159-1161.

Schleifer, K. H., & Kilpper-Bälz, R. (1984). Transfer of Streptococcus faecalis and Streptococcus faecium to the genus Enterococcus nom. rev. as Enterococcus faecalis

comb. nov. and Enterococcus faecium comb. nov. International Journal of Systematic and Evolutionary Microbiology, 34(1), 31-34.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics, 30(14), 2068-2069.

Seemann, T. (2015). Snippy: rapid haploid variant calling and core SNP phylogeny.

Seipke, R. F., Kaltenpoth, M., & Hutchings, M. I. (2012). Streptomyces as symbionts: an emerging and widespread theme? FEMS microbiology reviews, 36(4), 862-876.

Serio, A., Paparella, A., Chaves-Lopez, C., Corsetti, A., & Suzzi, G. (2007). Enterococcus populations in Pecorino Abruzzese cheese: biodiversity and safety aspects. Journal of food protection, 70(7), 1561-1568.

Sherman, J. M., Mauer, J. C., & Stark, P. (1937). Streptococcus fecalis. Journal of bacteriology, 33(3), 275.

Sifaoui, F., Arthur, M., Rice, L., & Gutmann, L. (2001). Role of penicillin-binding protein 5 in expression of ampicillin resistance and peptidoglycan structure in Enterococcus faecium. Antimicrobial agents and chemotherapy, 45(9), 2594-2597.

Singh, K. V., Coque, T. M., Weinstock, G. M., & Murray, B. E. (1998). In vivo testing of an Enterococcus faecalis efaA mutant and use of efaA homologs for species identification. FEMS Immunology & Medical Microbiology, 21(4), 323-331.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics, 30(9), 1312-1313.

Tavares, E. R., da Silva, L. F., Morey, A. T., de Oliveira, A. G., da Rocha, S. P. D., Ribeiro, R. A., ... & Yamada-Ogatta, S. F. (2019). Draft Genome Sequence of Vancomycin-Resistant Enterococcus faecium UEL170 (Sequence Type 412), Isolated from a Patient with Urinary Tract Infection in a Tertiary Hospital in Southern Brazil. Microbiol Resour Announc, 8(7), e01365-18.

Templer, K. J., Tay, C., & Chandrasekar, N. A. (2006). Motivational cultural intelligence, realistic job preview, realistic living conditions preview, and cross-cultural adjustment. Group & Organization Management, 31(1), 154-173.

Thiercelin, M. E., & Jouhaud, L. (1899). Sur un diplocoque saprophyte de l'intestin susceptible de devenir pathogene. CR Soc Biol, 5(26971.2).

Thorpe, H. A., Bayliss, S. C., Sheppard, S. K., & Feil, E. J. (2018). Piggy: a rapid, large-scale pan-genome analysis tool for intergenic regions in bacteria. Gigascience, 7(4), giy015.

Turner, K. M., Hanage, W. P., Fraser, C., Connor, T. R., & Spratt, B. G. (2007). Assessing the reliability of eBURST using simulated populations with known ancestry. BMC microbiology, 7(1), 30.

Ubeda, C., Taur, Y., Jenq, R. R., Equinda, M. J., Son, T., Samstein, M., ... & Pamer, E. G. (2010). Vancomycin-resistant Enterococcus domination of intestinal microbiota is enabled by antibiotic treatment in mice and precedes bloodstream invasion in humans. The Journal of clinical investigation, 120(12), 4332-4341.

van Schaik, W., Top, J., Riley, D. R., Boekhorst, J., Vrijenhoek, J. E., Schapendonk, C. M., ... & Willems, R. J. (2010). Pyrosequencing-based comparative genome analysis of the nosocomial pathogen Enterococcus faecium and identification of a large transferable pathogenicity island. BMC genomics, 11(1), 239.

Weigel, L. M., Clewell, D. B., Gill, S. R., Clark, N. C., McDougal, L. K., Flannagan, S. E., ... & Tenover, F. C. (2003). Genetic analysis of a high-level vancomycin-resistant isolate of Staphylococcus aureus. Science, 302(5650), 1569-1571.

Weinstein, M. P. (2019). Performance standards for antimicrobial susceptibility testing. Wayne, PA: Clinical and Laboratory Standards Institute

Werner, G., Fleige, C., Geringer, U., van Schaik, W., Klare, I., & Witte, W. (2011). IS element IS 16 as a molecular screening tool to identify hospital-associated strains of Enterococcus faecium. BMC infectious diseases, 11(1), 80.

Wickham, H. (2016). ggplot2: elegant graphics for data analysis. Springer.

Willems, R. J., Top, J., Marga van Santen, D., Coque, T. M., Baquero, F., Grundmann, H., & Bonten, M. J. (2005). Global spread of vancomycin-resistant Enterococcus faecium from distinct nosocomial genetic complex. Emerging infectious diseases, 11(6), 821.

Willems, R. J., Top, J., van Schaik, W., Leavis, H., Bonten, M., Sirén, J., ... & Corander, J. (2012). Restricted gene flow among hospital subpopulations of Enterococcus faecium. MBio, 3(4), e00151-12.

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., ... & Larsen, M. V. (2012). Identification of acquired antimicrobial resistance genes. Journal of antimicrobial chemotherapy, 67(11), 2640-2644.