5-2013

Molecular And Genomic Based Insights Into The Evolution Of Enterococcus Faecium From Commensal To Hospital-Adapted Pathogen

Jessica R. Galloway-Pena

MOLECULAR AND GENOMIC BASED INSIGHTS INTO THE EVOLUTION OF ENTEROCOCCUS FAECIUM FROM COMMENSAL TO HOSPITAL-ADAPTED PATHOGEN

by

Jessica Rhea Galloway-Peña, B.A.

APPROVED:

_____

Supervisory Professor, Barbara E. Murray, M.D.

_____

Steven Norris, Ph.D.

_____

Magnus Höök, Ph.D.

_____

Danielle Garsin, Ph.D.

_____

Richard Clark, Ph.D.

_____

Dean, The University of Texas Graduate School of Biomedical Sciences at Houston

MOLECULAR AND GENOMIC BASED INSIGHTS INTO THE EVOLUTION OF ENTEROCOCCUS FAECIUM FROM COMMENSAL TO HOSPITAL-ADAPTED PATHOGEN

A

DISSERTATION

Presented to the Faculty of the University of Texas Health Science Center at Houston

and

The University of Texas MD Anderson Center

Graduate School of Biomedical Sciences

in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

by

Jessica Rhea Galloway Peña, B.A.

Houston, TX

May, 2013

DEDICATION

I would like to dedicate my Doctor of Philosophy Dissertation to my Mother, Carol Michele Galloway, my Grandmother, Margaret Galloway, and my Aunt and Godmother, Patricia Lange. I would like to thank them for their dedication to my education and the betterment of my quality of life through knowledge and learning. I would also like to thank them for always supporting me in whatever I do, especially in my choices regarding my education.

ACKNOWLEDGEMENTS

I would first like to acknowledge all of the past and present members of Dr. Murray's and Dr. Arias' laboratories during the duration of my Ph.D. training for their feedback and help with my projects. I would like to give special thanks to include Agathe Bourgogne, Jouko Sillanpaa, Diana Panesso, and Kavindra Singh who were a great help to me by giving additional advice when I struggled with my experiments or needed to learn new techniques. I would also like to thank Kavindra Singh and Karen Jaques-Palaz for all their hard work with the immunizations and animal experiments, which contributed a great deal to the WxL project.

Secondly, I would like to acknowledge my past and present committee members for their feedback and encouragement on my projects, candidacy exam, and grant proposal. I would especially like to thank Magnus Höök for the use of his laboratory, equipment, and expert advice on protein purification and ECM binding, as well as his laboratory members Prasanth Francis, Cana Ross, and Xiaowen Liang who helped a great deal in teaching me how to purify proteins, characterize proteins, and contributions to the SPR, CD, and fluorescence data. Additionally, I would like to thank Dr. Hung Ton-That and his lab member, Julie Chang, for the use of the electron microscope and the immunogold electron micrographs taken for the WxL Project.

Third, I would like to give a special thank you to Dr. Steven Norris for his very special contribution to my graduate education. I thank him for being an amazing undergraduate research mentor, for discussing my options to further my education, for his encouragement to apply to graduate school, all of his recommendation letters throughout our professional relationship, and the honor of being a part of the MBID training grant.

And last but not least, I would like to thank my Ph.D. mentor Dr. Murray for all of her guidance and "tough love" throughout my graduate career. I would like to thank her for giving me creative freedom with my projects, letting me run with my ideas, supporting all of my grant and scholarship applications, trusting me with many responsibilities, and pushing my limits to make me a better scientist. I would also like to thank her for all of her help with my personal life and during the many struggles I had to overcome during my graduate career, for helping me by providing me emotional understanding and a home during divorce, family illness and death, and a pregnancy. Without you, I'm not sure I would have made it.

ABSTRACT

MOLECULAR AND GENOMIC BASED INSIGHTS INTO THE EVOLUTION OF ENTEROCOCCUS FAECIUM FROM COMMENSAL TO HOSPITAL-ADAPTED PATHOGEN

Publication No. _____

Jessica Rhea Galloway-Peña, B.A.

Supervisory Professor: Barbara E. Murray, M.D.

The basis for the recent transition of *Enterococcus faecium* from a primarily commensal organism to one of the leading causes of hospital-acquired infections in the United States is not yet understood. To address this, the first part of my project assessed isolates from early outbreaks in the USA and South America using sequence analysis, colony hybridizations, and minimal inhibitory concentrations (MICs) which showed clinical isolates possess virulence and antibiotic resistance determinants that are less abundant or lacking in community isolates. I also revealed that the level of ampicillin resistance increased over time in clinical strains. By sequencing the *pbp5* gene, I demonstrated an ~5% difference in the *pbp5* gene between strains with MICs <4ug/ml and those with MICs >4µg/ml, but no specific sequence changes correlated with increases in MICs within the latter group. A 3-10% nucleotide difference was also seen in three other genes analyzed, which suggested the existence of two distinct subpopulations of *E. faecium*. This led to the second part of my project analyzing concatenated core gene sequences, SNPs, the 16S rRNA, and phylogenetics of 21 *E. faecium* genomes confirming two distinct clades; a community-associated (CA) clade and hospital-associated (HA) clade. Molecular clock calculations indicate that these two clades likely diverged ~ 300,000 to $\geq$ 1 million years

ago, long before the modern antibiotic era. Genomic analysis also showed that, in addition to core genomic differences, HA *E. faecium* harbor specific accessory genetic elements that may confer selection advantages over CA *E. faecium*. The third part of my project discovered 6 *E. faecium* genes with the newly identified "WxL" domain. My analyses, using RT-PCR, western blots, patient sera, whole-cell ELISA, and immunogold electron microscopy, indicated that *E. faecium* WxL genes exist in operons, encode bacterial cell surface localized proteins, that WxL proteins are antigenic in humans, and are more exposed on the surface of clinical isolates versus community isolates (even though they are ubiquitous in both clades). ELISAs and BIAcore analyses also showed that proteins encoded by these operons bind several different host extracellular matrix proteins, as well as to each other, suggesting a novel cell-surface complex. In summary, my studies provide new insights into the evolution of *E. faecium* by showing that there are two distantly related clades; one being more successful in the hospital setting. My studies also identified operons encoding WxL proteins whose characteristics could also contribute to colonization and virulence within this species.

TABLE OF CONTENTS

I. Approval Sheet……i

II. Title Page……ii

III. Dedication ……...……iii

IV. Acknowledgments……iii

V. Abstract……v

VI. Table of Contents……...vii

VII. List of Illustrations……...xiv

VIII. List of Tables……..xvii

IX. Abbreviations……..xx

X. [1] Background……..1

[1.1] *Enterococcus Sp*……1

[1.2.2] Infections and Prevalence……..1

[1.2] Antibiotic Resistance in *Enterococcus faecium*……2

[1.2.1] Ampicillin Resistance……3

[1.2.2] Vancomycin Resistance……4

[1.3] Virulence Determinants of *Enterococcus faecium*……..5

[1.3.1] Enterococcal Surface Protein…….6

[1.3.2] Hyl-containing Plasmid…….6

[1.3.3] Collagen Binding Adhesin of *Enterococcus faecium* (Acm)……7

[1.3.4] Other MSCRAMMs that are Putative Adhesins and Pili……..8

[1.4] Population Biology of *Enterococcus faecium*……..8

[1.4.1] Origin/Host-specific Clonal Complexes……9

[1.4.2] Genome-based Insights into the Evolution of *E. faecium*……...10

[2] Chapter 1: Molecular Epidemiology, Clonality, and Antibiotic Resistance Analysis of Early *Enterococcus faecium* in the United States and South America……..12

[2.1] Chapter 1 Introduction……12

[2.2] Chapter 1 Materials and Methods……...15

[2.2.1] United States bacterial isolates and species-specific PCR…….……15

[2.2.2] South American specimen collection and testing……15

[2.2.3] Genomic DNA isolation, multi-locus sequence typing (MLST), and pulsed-field gel electrophoresis (PFGE)…….15

[2.2.4] Susceptibility testing……16

[2.2.5] Colony hybridization and *van* gene PCR……16

[2.2.6] Statistical analysis……17

[2.2.7] Bacterial isolates and *pbp5* gene sequencing……...18

[2.2.8] Other gene comparisons and phylogenetic studies……..18

[2.3] Chapter 1 Results……20

[2.3.1] Analysis of Clonality and Antibiotic Resistance among Early Clinical Isolates of *Enterococcus faecium* in the United States……20

[2.3.2] Molecular Epidemiology of Vancomycin-Resistant *Enterococcus faecium* in South American Hospitals……...…..35

[2.3.3] Analysis of PBP5 of Early United States *Enterococcus faecium* Isolates: Sequence Variation Alone Does Not Explain Increasing Ampicillin Resistance Over Time……38

[2.4] Chapter 1 Discussion……..52

[3] Chapter 2: Genomic Analyses Demonstrate a Distant Separation of the Hospital and Community Associated Clades of *E. faecium* ……56

[3.1] Chapter 2 Introduction…….....56

[3.2] Chapter 2 Materials and Methods……..59

[3.2.1] Selection of 100 orthologs……..59

[3.2.2] Comparative DNA sequence analysis……..59

[3.2.3] SNP analysis……65

[3.2.4] Molecular clock estimation…….65

[3.2.5] Analysis of the 16S rRNA and estimation of the time of divergence….65

[3.2.6] Genome characterization…….66

[3.2.7] Ortholog, phylogenetic and multi-locus sequence typing (MLST) analysis……66

[3.2.8] Identification of putative virulence-associated genes and antibiotic resistance determinants……67

[3.3] Chapter 2 Results……68

[3.3.1] Analysis of 100 genes from 21 *E. faecium* genomes illustrates the differences in core genome components of two clades……...68

[3.3.2] SNP analysis emphasizes clade-specific differences……...77

[3.3.3] Molecular clock determination using sSNP analysis shows an evolutionarily distant separation between the CA and HA clades……..81

[3.3.4] 16S rRNA and ribosomal protein analyses show clade-specific SNPs and a similar time of divergence to the SNP analysis……82

[3.3.5] Phylogenetic, multi-locus sequence typing (MLST) and gene content similarity analysis……87

[3.3.6] Mobile genetic elements……..87

[3.3.7] Distribution of genes encoding MSCRAMM-like proteins, putative virulence genes, antibiotic resistance determinants, and CRISPRs……96

[3.4] Chapter 2 Discussion……100

[4] Chapter 3: Defining the Role of "WxL" Proteins in *Enterococcus faecium*……..106

[4.1] Chapter 3 Introduction……..106

[4.2] Chapter 3 Materials and Methods…….109

[4.2.1] Strains, plasmids, and cultivation of bacteria……110

[4.2.2] Sequencing of loci and *in silico* genomic analysis……110

[4.2.3] Colony hybridizations…….....113

[4.2.4] RNA extraction and RT-PCR……114

[4.2.5] Construction of expression plasmids and purification of recombinant proteins……..116

[4.2.6] Fluorescence and Circular dichroism……118

[4.2.7] Production of polyclonal antibodies and purification of antigen-specific Igs……..119

[4.2.8] Whole-cell ELISA for detection of WxL proteins on the cell surface……...120

[4.2.9] Mutanolysin cell-wall extraction and western blot analysis……..120

[4.2.10] Immunogold Electron Microscopy……..121

[4.2.11] ELISA-type ligand-binding assays and Surface Plasmon Resonance (SPR)…….121

[4.2.12] Single-cross over insertion mutants and non-polar markerless deletion mutants……..117

[4.2.13] Whole-cell ligand binding assay, biofilm assay, and bile-salt assay……..125

[4.2.14] Detection of antibodies in patient sera and immunization/protection studies……126

[4.3] Chapter 3 Results……..128

[4.3.1] Six genes encoding WxL proteins in three different loci are found in *E. faecium* genomes……...130

[4.3.2] The phylogenetics and prevalence of WxL genes throughout *E. faecium* genomes…….136

[4.3.3] The predicted structure of the WxL proteins…….143

[4.3.4] The WxL proteins are expressed on the surface of bacterial cells although they are differentially exposed depending on the clade of the strain……....148

[4.3.5] Recombinant proteins from the WxL Locus A bind a variety of extracellular matrix proteins……..152

[4.3.6] DufA shows self-association and association with SwpA and LwpA…….162

[4.3.7] WxL Locus A and WxL Locus C mutants are attenuated in type I collagen binding, but not type V collagen binding or biofilm formation……..164

[4.3.8] The WxL Loci could be involved in bile and salt stress response…….159

[4.3.9] WxL proteins are antigenic, and may have use in protection against *E. faecium* infections…….167

[4.4] Chapter 3 Discussion……173

[5] Concluding Remarks…….....180

XI. Bibliography……185

XII. Vita……206

LIST OF ILLUSTRATIONS

FIGURE 1. Key events that occurred in the early clinical United States isolates of this study.....30

FIGURE 2. A comparison between early United States CC17 isolates and non-CC17 isolates for the presence of ampicillin resistance, vancomycin resistance, the esp_{Efm} gene, and the hyl_{Efm} gene……..33

FIGURE 3. A comparison between CC17 and non-CC17 isolates for the presence of the putative adhesins and pili (*fms)* genes in early United States *E. faecium* isolates….…….34

FIGURE 4. Frequency of putative virulence genes in 35 VR *E. faecium* isolates from the Andean region of South America PBP5 phylogeny coincides with MIC……...37

FIGURE 5. PBP5 phylogeny coincides with MIC**.** ……48

FIGURE 6. UPGMA trees of four *E. faecium* genes shows isolates generally fall into two distinct lineages ……..50

FIGURE 7. Position of the 100 genes analyzed on the TX16 (DO) chromosome……..64

FIGURE 8. UPGMA phylogenetic tree of the concatenated 100 gene sequence (106,818bp)…….71

FIGURE 9. UPGMA phylogenetic tree of the concatenated amino acid sequence (35,606bp)……...76

FIGURE 10. UPGMA phylogenetic tree of the concatenated SNP sequence (5,392bp)……80

FIGURE 11. An UPGMA tree representing the evolutionary relationship of the strains using the 16S rRNA nucleotide sequence for each of the 21 *E. faecium* strains……..83

FIGURE 12. An UPGMA tree representing the evolutionary relationship of the strains using the *rpmF* nucleotide sequence for each of the 21 *E. faecium* strains……..85

FIGURE 13. *Enterococcus faecium* phylogenetics……91

FIGURE 14. The projected evolution of the two clades of *E. faecium*……104

FIGURE 15. The WxL domain of 6 WxL proteins in *E. faecium*……131

FIGURE 16. The organization of the 3 putative operons containing WxL genes in *Enterococcus faecium*……..132

FIGURE 17. RT-PCR of intergenic regions of TX1330……..133

FIGURE 18. RT-PCR of intergenic regions of TX82……..134

FIGURE 19. UPGMA phylogenetic trees and identities of WxL loci genes throughout *E. faecium* genomes……..138

FIGURE 20. Variations upstream and downstream of the WxL loci……..141

FIGURE 21. Colony hybridizations for the genes encoding WxL proteins show that these genes are ubiquitous in *E. faecium* strains, and do not depend on origin……..142

FIGURE 22. The UV absorbance spectra of purified TX82 WxL Locus A recombinant proteins……144

FIGURE 23. Circular dichromism spectra of purified TX82 WxL Locus A recombinant proteins……....146

FIGURE 24. Steady-state intrinsic tryptophan fluorescence spectra show differences between the monomer and dimer forms of SwpA as well as differences under different salt concentrations for SwpA (dimer), DufA, and LwpA……147

FIGURE 25. Western blots for WxL proteins of Locus A and C of mutanolysin cell wall extracts from TX82 cells……..149

FIGURE 26. Immunogold electron microscopy shows the WxL Locus C proteins on the surface of TX82 cells……150

FIGURE 27. The differences in surface display/expression of WxL proteins between community clade isolates and hospital clade isolates…….151

FIGURE 28. Binding of extracelluar matrix proteins by recombinant proteins from WxL Locus A……..152

FIGURE 29. An initial screen using Surface Plasmon Resonance shows binding responses for Locus A recombinant proteins to different immobilized ECM……...154

FIGURE 30. Surface Plasmon Resonance shows interactions between Type I collagen, DufA, and SwpA…….……156

FIGURE 31. Surface Plasmon Resonance shows interactions between fibronectin, DufA, SwpA, and LwpA…….157

FIGURE 32. Surface Plasmon Resonance shows specific interactions between of DufA and SwpA with the Hep2-40K chain of fibronectin……....159

FIGURE 33. Surface Plasmon Resonance shows specific interactions between of DufA with itself, SwpA and Lwp A that is dependent on salt concentrations…...…….163

FIGURE 34. The WxL Locus A and WxL Locus C deletion mutants are attenuated in binding to type I collagen……...164

FIGURE 35. The WxL Locus A and C mutants are not attenuated in biofilm formation.…..165

FIGURE 36. The Locus B WxL mutant survives better than the wild-type after bile salt exposure……166

FIGURE 37. The WxL proteins of Locus A and Locus C are antigenic in patients……169

FIGURE 38. Protection against TX82 endocarditis after active immunization with LwpA and LwpC…….……..…….171

LIST OF TABLES

TABLE 1. Primers used in this study for colony hybridization of putative MSCRAMMs and virulence factors 16

TABLE 2. Primers used in this study to sequence the complete *pbp5* gene of *Enterococcus faecium*…….19

TABLE 3. Relevant background and characteristics of the early clinical U.S. *E. faecium* isolates included in this study…….....24

TABLE 4. Results for early clinical U.S. isolate colony hybridizations.....32

TABLE 5. Molecular typing of the vancomycin resistant *E. faecium* isolates collected from the Andean region of South America……36

TABLE 6. Single nucleotide polymorphisms at each nucleotide position of the consensus sequence of *pbp5* in the 29 U.S. strains analyzed……38

TABLE 7. Analysis of silent polymorphisms and the codon usage differences between *pbp5*-S, hybrid, and *pbp5*-R strains……...42

TABLE 8. Amino Acid Changes Between Ampicillin Susceptible and Resistant *Enterococcus faecium* Stains from the United States…….....44

TABLE 9. The sequence type, country, date, and source of isolation for the 21 sequenced *E. faecium* genomes used in this study……60

TABLE 10. 100 core genes chosen for analysis and concatenation in this study……...61

TABLE 11. Percent identity and divergence score matrix of the 100 concatenated gene nucleotide sequence……68

TABLE 12. Clade analysis of 91 individual genes……73

TABLE 13. Percent identity and divergence score matrix of the concatenated amino acid sequence…….73

TABLE 14. Percent identity and divergence score matrix of the concatenated SNP sequence…….77

TABLE 15. Molecular clock/time of divergence estimates based on sSNP analysis within and between clades……81

TABLE 16. Hospital-associated clade unique mobile elements (IS elements/ transposons).....88

TABLE 17. Presence of genes encoding MSCRAMMs and pilins among 21 *E. faecium* genomes……...93

TABLE 18. Antibiotic resistance gene profiles of the 21 *E. faecium* strains……..98

TABLE 19. Summary of CRISPRs found in *E. faecium* sequenced strains……99

TABLE 20. Bacterial strains and plasmid used in this study……...109

TABLE 21. Primers used for the re-sequencing of the WxL loci…….111

TABLE 22. Primers for probes used in hybridizations of genes encoding WxL proteins. ……113

TABLE 23. Primers used for RT-PCR of the intergenic regions of WxL loci……..115

TABLE 24. Primers used to make WxL recombinant proteins……..118

TABLE 25. Primers used for single cross over insertion mutants in TX1330……...123

TABLE 26. Primers used for markerless non-polar deletions on the WxL Loci in TX82….124

TABLE 27. Summary of Locus A recombinant protein characteristics……..144

TABLE 28. Secondary-structure content of WxL proteins determined by CD……..145

TABLE 29. Surface Plasmon Resonance analysis shows recombinant proteins from Locus A bind host ECM proteins……155
TABLE 30. Surface Plasmon Resonance analysis shows recombinant proteins from Locus A bind each other……162

ABBREVIATIONS

AFLP: Amplified Fragment Length Polymorphism

Amp^R: Ampicillin Resistant

AREfm: Ampicillin Resistant *E. faecium*

CA: Community Associated

CC17: Clonal Complex 17

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats

FCA: Freund's Complete Adjuvant

FICA: Fruend's Incomplete Adjuvant

FnBP: Fibronectin Binding Protein

Gen^R: Gentamicin Resistant

HA: Hospital Associated

HLR: High Level Resistance

MICs: Minimal Inhibitory Concentrations

MLST: Multilocus Sequence Typing

MSCRAMMs: Microbial Surface Components Recognizing Adhesive Matrix Molecules

PAI: Pathogenicity Island

PFGE: Pulsed-field Gel Electrophoresis

PBPs : Penicillin-binding Proteins

SPR: Surface Plasmon Resonance

ST: Sequence Type

STR^{HLR}: Streptomycin High Level Resistance

UPGMA: Unweighted Pair Group Method with Arithmetic Mean

Van^R: Vancomycin Resistant

VREfm: Vancomycin Resistant *Enterococcus faecium*

[1] BACKGROUND

[1.1] *Enterococcus* spp.

Enterococci were previously classified as group D streptococci via the Lancefield serological typing system (Murray, 1997). In the 1980's, enterococci were removed from the genus *Streptococcus* and a new genus, *Enterococcus*, was created based on 16S rRNA and DNA-rDNA hybridizations (Arias and Murray, 2008; Klein, 2003). The genus *Enterococcus* consists of over 30 species of facultatively anaerobic, low G+C content, lactic acid producing, Gram-positive cocci (Fisher and Phillips, 2009; Willems and van Schaik, 2009). *Enterococcus* species have historically been used widely in the food industry and as probiotics, but over the past 30 years have become the third most common organisms recovered from the nosocomial environment (Foulquie Moreno et al., 2006; Hidron et al., 2008). *Enterococcus* species can be found in environmental, animal, insect, nematode and human sources, existing as an intrinsic part of both animal and human microflora. The most common two species found in the human gastrointestinal tract, as well as the two predominate species that cause infections, are *Enterococcus faecalis* and *Enterococcus faecium* (Fisher and Phillips, 2009; Hidron et al., 2008).

[1.1.2] Infections and Prevalence

The most common infections caused by enterococci are urinary tract infections, bacteremia, endocarditis, soft tissue or wound infections, and peritonitis/intra-abdominal infections (Murray, 1990). Although, *E. faecalis* has historically accounted for 80-90% of all enterococcal clinical isolates, the most recent data indicates that *E. faecium* currently accounts for approximately 30-40% of all enterococcal clinical isolates (Hidron et al., 2008; Murray,

1997). Thirty years ago, *E. faecium* was rarely encountered in human infections and considered a harmless commensal. However, reports of outbreaks of *E. faecium* began to be reported in the United States in the 1980's, and now are a predominant cause of hospital-acquired infections (Bush et al., 1989; Coudron et al., 1984; Eliopoulos et al., 1988; Grayson et al., 1991; Sapico et al., 1989). This dissertation serves to propose reasons for the transition of *E. faecium* from a harmless commensal to a formidable nosocomial pathogen through antibiotic resistance, particular virulence determinants, and a genomic background difference separating commensal and clinical isolates into two distinct clades.

[1.2] Antibiotic Resistance in *Enterococcus faecium*

Currently, 80% of all clinical *E. faecium* recovered in the United States are resistant to vancomycin, and an even higher 90% are resistant to ampicillin (Hidron et al., 2008). In addition to resistance to the two predominant therapies against enterococcal infections, *E. faecium* has a number of intrinsic resistances, such as to cephalosporins, clindamycin, and low-level aminoglycoside resistance. Moreover, *E. faecium* has a number of acquired resistances that are thought to emerge via plasmids and transposons from other organisms. As part of the normal gastrointestinal flora, these organisms are therefore exposed to antibiotic selective pressure when any infections are treated. Such antibiotics include ciprofloxacin, tetracycline, chloramphenicol, trimethoprim, and high-level resistance to aminoglycosides. All of these scenarios together make *E. faecium* infections incredibly difficult to treat via monotherapy, and almost eliminates the chance for synergistic bactericidal activity (Murray, 1997). More recently, *E. faecium* has also been reported to be resistant to more recently available antibiotics such as linezolid and daptomycin (Arias and Murray, 2008).

[1.2.1] Ampicillin Resistance

A study analyzing the progression of ampicillin resistance in *E. faecium* was first reported in a 22-year study using isolates from one hospital in Boston (Grayson et al., 1991). Although it is well known that ampicillin minimal inhibitory concentrations (MICs) of clinical *E. faecium* are much higher than they were 30 years ago, there has been no study that includes isolates from the earliest outbreaks of ampicillin resistant *E. faecium* from all over the United States.

It has long been thought that the expression of a low affinity class B penicillin-binding protein (PBP5) is involved in decreased ampicillin susceptibility (Fontana et al., 1985; Williamson et al., 1985). In general, penicillin-binding proteins are involved in peptidoglycan polymerization. Class B penicillin binding proteins (PBPs) are specifically D-D-transpeptidases that cooperate with glycosyltransferases to synthesize peptidoglycan (Zapun et al., 2008). PBP5 has low affinity for penicillins and can assume the transpeptidase function when other, more penicillin-susceptible PBPs are inhibited by a beta-lactam, leading to this species' intrinsic low-level resistance (Fontana et al., 1985). In some isolates, it seems that intermediate levels of resistance occur from increased expression of PBP5 (Fontana et al., 1994; Klare et al., 1992), whereas some clinical strains exhibiting higher MICs contain a very low affinity PBP5 (at least 10-times lower affinity) (Rybkine et al., 1998; Zorzi et al., 1996). Moreover, it has been shown that the gene encoding the low-affinity PBP5 can be transferred horizontally, suggesting the possibility that the low-affinity PBP5 can be circulated between clinical isolates (Rice et al., 2005).

Several amino acid changes (alone or in combination) in PBP5 have been correlated with its decreased affinity for β-lactams, although these have differed from study to study

(Ligozzi et al., 1996; Rybkine et al., 1998; Zorzi et al., 1996). For many of these correlation-based studies, it is not known if the strains were even somewhat related. Furthermore, in studies based on mutations in isogenic strains, the mutations resulted in only modest increases in MIC, and thus do not explain the full extent of resistance (Rice et al., 2004; Sifaoui et al., 2001; Zorzi et al., 1996). Therefore, whether specific amino acid variations can be correlated with increasing MICs displayed in clonally related clinical strains has yet to be determined.

[1.2.2] Vancomycin Resistance

Due to the increased prevalence of ampicillin resistance and high-level resistance to aminoglycosides in *E. faecium* in the 1980's, glycopeptides such as vancomycin represented one of the last therapeutic options during that time (Bonten et al., 2001). The first vancomycin resistant isolate was isolated in Europe in 1986 (Uttley et al., 1988). Vancomycin resistant *E. faecium* (VREfm) was reported in the United States soon thereafter, where outbreaks have continued since (although outbreaks of VREfm were delayed in Europe until relatively recently) (Arias and Murray, 2012). The basic mechanism of action of vancomycin is that it disrupts cell wall synthesis by binding to pentapeptide peptidoglycan precursors, specifically the terminal dipeptide D-ala-D-ala (Murray, 1997). Enterococci become vancomycin-resistant by generating pentapeptide precursors with alternate termini, either D-ala-D-lactate (Van A, B ,and D) or D-ala-D-serine (VanC, E, G, and L). A number of Van gene clusters have been identified that are responsible for this mechanism of resistance. The *vanA* gene cluster is the one most commonly found in clinical isolates, and is typically associated with the transposon Tn*1546* (Arias and Murray, 2008; Bonten et al., 2001; Murray, 1997).

Although VREfm have been reported in many countries all over the world, studies over the past several years have shown that the epidemiology of VREfm is quite different between

the United States and Europe. In the United States, colonization and infections by VREfm are associated with hospitals and correlate with increased antibiotic usage. It is thought that the epidemiology in United States hospitals is influenced by several clinically-associated variables (such as antibiotic pressure and environmental contamination), leading to patient colonization (Bonten et al., 2001).

In contrast, VREfm infections are not as prevalent in the hospitals in Europe as they are in the United States. However, there seems to be a community and animal reservoir of VREfm in Europe that is lacking in the United States. So, although VREfm is an uncommon cause of nosocomial infections in most parts of Europe until recently, colonization with VREfm in healthy individuals and animals is common. Some scientists explain this phenomenon with the use of the glycopeptide avoparcin in animal husbandry and agriculture in Europe (Bonten et al., 2001) Thus, although vancomycin resistant isolates exist in the community in Europe, these are not often causes of nosocomial infections. In contrast, United States VREfm are important nosocomial pathogens that commonly colonize hospitalized patients, but these isolates do not come from a community reservoir (Bonten et al., 2001). This observation begs the question: what are the differences in these two groups of isolates?

[1.3] Virulence Determinants of *Enterococcus faecium*

To date, many of the potential virulence determinants that have been described in *E. faecium* are members of a class of surface proteins called MSCRAMMs (microbial surface components recognizing adhesive matrix molecules), which are typically adhesins and pili. MSCRAMM–like proteins include a secretion signal sequence, a region with immunoglobulin-like folds that is usually the ligand-binding domain, and a C-terminal LPxTG-like motif that is recognized and cleaved by sortases (transpeptidase enzymes) which covalently attach the

protein to the cell-wall peptidoglycan (Hendrickx et al., 2009a; Hendrickx et al., 2007; Sillanpaa et al., 2008). It is thought that these proteins contribute to virulence by conferring attachment to host tissues and increase the colonizing ability of the strains that contain them (Hendrickx et al., 2009a; Sillanpaa et al., 2009a).

[1.3.1] Enterococcal Surface Protein

The *esp* gene of *E. faecium* (*espEfm*), encoding an enterococcal surface protein, has been shown to play a role in biofilm formation, experimental urinary tract infection, and endocarditis (Heikens et al., 2007; Leendertse et al., 2009; Heikens et al., 2011). The *esp* gene encodes an MSCRAMM-like protein and is located on a putative pathogenicity island (Leavis et al., 2004). Expression of *esp* has been associated with initial adherence and biofilm formation on abiotic surfaces, because insertional mutants were attenuated in these processes (Heikens et al., 2007). In addition, infections with *esp*-positive cells were shown to transiently aggravate urinary tract infections in mice (Leendertse et al., 2009) and an *esp*-deficient mutant was attenuated in experimental endocarditis (Heikens et al., 2011). However, isogenic mutants failed to demonstrate a role for Esp in peritonitis (Leendertse et al., 2009) and GI colonization (Heikens et al., 2009) .

[1.3.2] Hyl-containing Plasmid

The *hyl* gene of *E. faecium* (hyl_{Efm} which encodes a protein with homology to a family 84 glycosyl hydrolase) was postulated as a potential virulence factor due to the observation that it was found in some VREfm clinical strains, but not found in strains from community origins (fecal samples from healthy volunteers, animals, probiotics, and sewage) (Rice et al., 2003). The *hyl*-like gene was shown to be on a plasmid which increased the ability of strains to colonize the murine gut (Rice et al., 2009). Another study also showed that the hyl_{Efm} gene was

carried on large plasmids which were readily transferred from clinical strains to recipient *E. faecium* strain, possibly due to the presence of *pilA* (*fms21*) (Kim et al., 2010). The conjugative transfer of one *hyl*-containing plasmid conferred resistance to vancomycin (*vanA*) as well as increased lethality in the mouse peritonitis model (Arias et al., 2009). However, it was determined that the *hyl* gene, itself, does not contribute to peritonitis and the other gene(s) located on this plasmid possibly responsible for the increased virulence and colonization are currently being determined (Panesso et al., 2011).

[1.3.3] Collagen binding adhesin of *Enterococcus faecium* (Acm)

Acm (adhesin of collagen from *E. faecium*) is an MSCRAMM that interacts primarily with collagen type I, and secondarily with collagen type IV (Nallapareddy et al., 2008a; Nallapareddy et al., 2008b; Nallapareddy et al., 2003). Although *acm* is present in almost all strains isolated, it has been shown that a functional Acm is expressed exclusively by those with clinical origins. On the other hand, pseudogenes were often found in strains from community origins, where a premature stop codon disrupts collagen binding (Nallapareddy et al., 2008b). In a mixed infection endocarditis model, it was shown that the vegetations on the heart valves consisted of approximately 3 times as many wild-type cells as *acm* deletion mutant cells and that the *acm* deletion mutant was attenuated in initial adherence and colonization of the heart valves (Nallapareddy et al., 2008a). In addition, sera from patients with endocarditis contain anti-Acm IgG antibodies (Nallapareddy et al., 2008a). All this evidence together has shown that Acm likely plays a role in pathogenesis, namely in endocarditis, and elicits an antibody response.

[1.3.4] Other MSCRAMMs that are Putative Adhesins and Pili

Analysis of the draft genome of TX16 revealed the presence of 15 genes, called *fms* genes (*E. faecium* surface proteins), encoding proteins with MSCRAMM-like qualities thought to be putative adhesins or pili. Eleven of these 15 MSCRAMMs clustered into four pilus-like gene clusters, each with an associated class C sortase. Nine of the eleven showed homology to the Ebp pili of *E. faecalis* and contained pilus assembly motifs (Sillanpaa et al., 2008). Two of these, encoded by *fms9* (*pilB*) and *fms21* (*pilA*), have been shown to polymerize into a high-molecular-mass protein bands in western blots and to polymerize into pili at the cell surface using electron microscopy (Hendrickx et al., 2008; Sillanpaa et al., 2008). Three of the predicted pili showing homology to Ebp pili in *E. faecalis* (encoded by *empABCfm* in *E. faecium*) have been shown to be involved in biofilm formation and virulence in the murine model of urinary tract infections, and endocarditis (Sillanpaa et al., 2010). In addition, Scm (second collagen adhesin of *E. faecium*), also known as Fms10, has been shown to bind primarily to collagen type V, and less so to collagen type I and fibrinogen (Sillanpaa et al., 2008).

[1.4] Population Biology of *Enterococcus faecium*

With *E. faecium* outbreaks reported on 5 continents and seemingly different characteristics in *E. faecium* between the United States and Europe (especially in regards to antibiotic resistance as discussed previously), there was an urgency to adapt molecular typing methods in order to understand the epidemiology of these hospital outbreaks. One of the first molecular typing methods was pulsed-field gel electrophoresis (PFGE), which is still considered the "gold standard" for short term hospital epidemiology (Murray, 1990). In order to study the genetic relatedness between strains from different origins (human vs. animal),

amplified fragment length polymorphism analysis (AFLP) was used (Willems et al., 2000). In 2002, a multi-locus sequence typing (MLST) scheme was proposed for *E. faecium*, as had been developed for 30+ organisms at the time, to determine global population biology (Homan et al., 2002). These three methodologies have given us much of the information we have about the population structure of *E. faecium* to date.

[1.4.1] Origin/Host-specific Clonal Complexes

Although PFGE and AFLP gave scientists their first insight into the population structure of *E. faecium*, showing different clones and host specific genetic lineages, both methods were unsuitable for reproducible data exchange worldwide (Jureen et al., 2004a; Werner et al., 2003; Willems et al., 2000). The first studies using MLST of *E. faecium* confirmed the genogroups previously found using AFLP, including a hospital-related genogroup (Homan et al., 2002). Subsequent studies using MLST and the eBURST algorithm on *E. faecium* isolates from multiple sources revealed that hospital outbreak isolates clustered to a sub-population designated C1, which was renamed clonal complex 17 based on the thought that the founder of this lineage was ST17 (Leavis et al., 2006a; Top et al., 2007; Top et al., 2008; Willems et al., 2005). Further studies showed a strong correlation between the clonal complex 17 (CC17) genetic lineage and ampicillin resistance, quinolone resistance, the pathogenicity island (PAI) containing *esp*, the *hyl* gene, and a number of surface proteins containing the LPxTG-motif (Hendrickx et al., 2007; Leavis et al., 2006a; Leavis et al., 2006b; Nallapareddy et al., 2008b; Rice et al., 2003; Sillanpaa et al., 2009a; Willems et al., 2001). Based on these studies, it was predicted that CC17 was responsible for the worldwide emergence of nosocomial VREfm (Leavis et al., 2006a).

It was soon found, however, that inferring evolutionary descent using eBURST is unreliable for organisms that have high rates of population recombination, such as *E. faecium* (Turner et al., 2007). Although eBURST is no longer generally accepted to determine ancestry, MLST is still an acceptable typing method for *E. faecium*. It still stands that clinically-derived isolates are of particular sequence types (STs) (ST17, or single or double locus variants of ST17) which are not found in community isolates (Didelot and Falush, 2007; Willems et al., 2011; Willems and van Schaik, 2009). Other studies using different modeling such as neighbor-net trees (CLONALFRAME) and Bayesian modeling (BAPS), confirmed the suspicions of Turner et al. by showing that major hospital clones that were previously linked by eBURST grouped separately (Didelot and Falush, 2007; Tang et al., 2009), nullifying the idea that *E. faecium* hospital-associated isolates evolved recently from a single common ancestor that was ST17.

[1.4.2] Genome-based Insights into the Evolution of *E. faecium*

Comparative genomic hybridization using isolates from diverse backgrounds in a mixed whole genome array also supported the existence of a specific hospital-specific subpopulation. CC17 isolates overlapped with this phylogenomic group, as well as non-CC17 isolates from clinical origins, demonstrating that horizontal transfer of genes involved in colonization and infection is not limited to the CC17-related genogroup. It was also proposed in this work that ~150 genes, transposases and IS elements, specifically IS*16*, were enriched in hospital isolates, possibly conferring genomic plasticity to these strains (Leavis et al., 2007). This was confirmed by another study that showed IS*16* was exclusively found in hospital-associated isolates of *E. faecium*; specifically CC17-related strains (Werner et al., 2011).

A study using pyrosequencing of 7 draft genomes determined a genome-based phylogeny based on ~650 orthologs conserved in different strains (van Schaik et al., 2010; van Schaik and Willems, 2010). Although this study suggested that human infection isolates were different from human commensal isolates based on conserved genes (not accessory genomes), there was only one isolate from the commensal group, making these data unreliable. However, this study did show as high as a 12% difference in gene content between two closely related strains by conserved genome analysis, highlighting the importance of the gain and loss of mobile genetic elements (van Schaik et al., 2010; van Schaik and Willems, 2010).

Approximately 30+ *E. faecium* strains have been sequenced, partially assembled, and annotated (Palmer et al., 2010). Only one other genome (the *vanB* ST17 Aus0004) besides the TX16 (DO) genome discussed in this dissertation has been fully assembled and published (Lam et al., 2012). So, although partially assembled draft genomes have given us important insights into the evolution of *E. faecium*, the availability of two complete and closed genomes is more advantageous to compare and understand the genetic features that may explain the emergence of this species as a nosocomial pathogen.

[2] CHAPTER 1

Molecular Epidemiology, Clonality, and Antibiotic Resistance Analysis of Early *Enterococcus faecium* in the United States and South America

[2.1] CHAPTER 1 INTRODUCTION

The emergence of multi-resistant *E. faecium* world-wide and hospital outbreaks on several continents were previously attributed to a particular genetic lineage designated CC17 (Leavis et al., 2006a). However, recent publications (Turner et al., 2007) suggest that clustering of MLST-based allelic profiles using eBURST is unreliable for phylogenetic prediction of *E. faecium* due to its relatively high recombination rates. Nevertheless, certain STs (ST16, ST17, ST18, and ST19) are prominent among hospital- associated isolates, and are typically associated with CC17. CC17 genogroup isolates are characterized by ampicillin resistance and frequently resistant to vancomycin. CC17 strains also have higher rates than non-CC17 isolates of putative virulence determinants and colonization factors such as the hyl_{Efm}-containing plasmid, esp_{Efm}, and genes encoding MSCRAMMs (Hendrickx et al., 2007; Leavis et al., 2004; Rice et al., 2003; Sillanpaa et al., 2008; Sillanpaa et al., 2009a; Willems et al., 2001).

Although outbreaks of ampicillin-resistant *E. faecium* in the United States were first reported in the early 1980's (Bush et al., 1989; Coudron et al., 1984; Sapico et al., 1989), clonality and molecular genetics of early ampicillin resistant *E. faecium* (AREfm) and VREfm United States isolates had not been characterized. Similarly, in South America, VREfm infections and isolates with CC17 related STs have been reported since 1998 (Dalla Costa et

al., 1998; Marin et al., 1998), yet there are limited data regarding the population genetics or presence of the potential virulence determinants in *E. faecium* in South America.

This chapter discusses three related published studies. In the first study (Galloway-Pena et al., 2009), 53 early clinical isolates collected in the United States between 1971 and 1994 were analyzed for the MLST, the presence of putative virulence genes, and their resistance to ampicillin, vancomycin, gentamicin and streptomycin. Among these isolates, 16 different STs were identified in 9 different regions of the United States. It was found that the earliest CC17 related isolates have been circulating since 1982. Characteristics of CC17 isolates over this time period included increasing ampicillin resistance, the increased presence of hyl_{Efm} and esp_{Efm}, the emergence of vancomycin resistance, and the presence of multiple *fms* genes.

The second study, which includes selected data from a prospective surveillance study (Panesso et al., 2010) using isolates from 32 hospitals and four countries, is included with the aim of characterizing the molecular genetics of enterococci circulating in northern South America. Genotyping was performed on the 35 VREfm isolates (out of the 760 enterococci recovered) using PFGE and MLST. These isolates were also evaluated for the presence of the same 16 putative virulence genes assessed in the previous study (*fms* genes, *espEfm*, and *hylEfm*). Eleven pulsotypes were found among the VREfm isolates, and most isolates belonged to sequence types 412 and 18. The *fms5-6-9* and *fms11-19-16* clusters were detected in all VREfm isolates isolated from this region. However, *espEfm* and *hylEfm* were detected in only 69% and 23% of the isolates, respectively. Therefore, the recent population genetics of VREfm in South America appear to parallel those of the early U.S, isolates in the first study. These data suggest that more CC17-related *E. faecium* may disseminate in this region in the coming years, and result in multiple outbreaks as they have in the United States.

The third study reflects back on the first study which showed that ampicillin resistance has increased steadily over the past three decades within United States (Galloway-Pena et al., 2011). This study found that out of 41 ampicillin resistant isolates analyzed, 9 isolates with MICs < 4µg/ml have a PBP5 consensus sequence that is distinctly different from the PBP5 consensus of the 32 isolates with MICs > 4µg/ml, with ~5% difference between these. However, we were not able to definitively determine consistent amino acid changes that correlated with specific increases in the MICs of ampicillin. After analysis of *pbp5* and three other genes encoding cell wall/surface proteins to determine if this 5% difference was specific to *pbp5*, it was determined that there were two distinct evolutionary groups for each gene.

[2.2] CHAPTER 1 MATERIALS AND METHODS

[2.2.1] United States bacterial isolates and species-specific polymerase chain reaction.

A total of 53 *E. faecium* isolates were selected from our collection of >450 *E. faecium* clinical (~75% of the collection) and nonclinical isolates (~25% of the collection; including community, fecal, and animal isolates) collected between 1970 and the present from diverse origins and geographic locations with >15 countries represented. To confirm the identity of the isolates, species-specific amplification of the $_{D}$-alanyl-$_{D}$-alanine ligase gene (*ddl*) was used (Dutka-Malen et al., 1995) in addition to initial biochemical tests. The isolates were selected based on a set of pre-determined criteria which included: (1) the earliest United States *E. faecium* isolates in our collection; (2) isolates recovered from clinical samples, with the majority representing nosocomial outbreaks (Bush et al., 1989; Coudron et al., 1984; Eliopoulos et al., 1988; Grayson et al., 1991; Handwerger et al., 1993; Sapico et al., 1989); and (3) diversity in geographical origin, with 9 different areas in the United States represented (see section 2.3.1 Table 3). (Galloway-Pena et al., 2009)

[2.2.2] South American specimen collection and testing.

A multicenter study was performed from February 2006 through February 2008 to evaluate the molecular epidemiology of enterococci in four South American countries. Each participating tertiary-care hospital (a total of 32 hospitals, from the Latin American Network of Antimicrobial Resistance) in Colombia (22 hospitals in six cities), Ecuador (5 hospitals in one city), Peru´ (3 hospitals in one city), and Venezuela (2 hospitals in one city) collected consecutive enterococcal isolates (duplicate organisms from the same patient were excluded). The samples were collected by the clinical laboratory from each participating hospital and corresponded to samples from hospitalized patients in Colombia, Peru´, and Venezuela. In Ecuador, the clinical laboratory also collected samples from ambulatory services and outpatient clinics. The isolates originated from the following clinical specimens: blood, urine, secretions from surgical wounds, peritoneal fluid, abdominal abscesses, joint aspirates, osteomyelitis aspirates, pleural fluid, pericardial effusion, cerebral abscesses, and cerebrospinal fluid (CSF). In an attempt to avoid isolates that likely represented colonization, enterococci recovered from sputum, rectal swabs, catheters, or skin (unless they originated in an infected surgical wound) were excluded. Each hospital identified the microorganisms by using either automated methods (performed with the Vitek or MicroScan system) or manual methods. (Panesso et al., 2010)

[2.2.3] Genomic DNA isolation, multi-locus sequence typing, and pulsed-field gel electrophoresis.

With the use of the DNeasy Blood and Tissue Kit (Qiagen), genomic DNA was extracted from 5-mL cultures incubated overnight in brain-heart infusion broth (Difco Laboratories). Polymerase chain reaction (PCR) and multilocus sequence

typing (MLST) were performed as described elsewhere (Homan et al., 2002). Fragments of 7 housekeeping genes (*atpA*, *ddl*, *gdh*, *purk*, *gyd*, *pstS*, and *adk*) were sequenced, allelic profiles were obtained, and a sequence type (ST) was designated for each unique allelic profile based on the MLST website (http://efaecium.mlst.net). For select isolates that had a pulsotype identical to that of typed strains recovered from the same area or during the same outbreak, only *purK* was sequenced, and an inferred sequence type was given. Pulsed-field gel electrophoresis (PFGE) was performed as described elsewhere (Nallapareddy et al., 2008b). (Galloway-Pena et al., 2009)

[2.2.4] Susceptibility testing.

Susceptibilities to ampicillin and vancomycin and high-level resistance to gentamicin and streptomycin were determined using standard agar dilution methods. In accordance with Clinical and Laboratory Standards Institute guidelines, minimal inhibitory concentrations (MICs) of 16 mg/mL for ampicillin and 32 mg/mL for vancomycin were considered to be resistant. Isolates for which growth on plates occurred at concentrations of 500 mg/mL gentamicin and 2000 mg/mL streptomycin were considered to have high-level resistance. (Galloway-Pena et al., 2009)

[2.2.5] Colony hybridization and *van* gene PCR.

Preparation of colony lysates on nylon membranes and hybridization under high-stringency conditions were performed as described elsewhere (Singh et al., 1998). DNA probes for hyl_{Efm}, esp_{Efm}, and the 14 *fms* genes were made using primers published elsewhere (Rice et al., 2003; Sillanpaa et al., 2009b) and found in Table 1. Of note, all isolates in the present study had the *acm* gene (*fms8*) that encodes a collagen-binding adhesin, consistent with our previous finding of its presence in the vast majority of clinical isolates (Nallapareddy et al., 2008b). These probes were radiolabelled using the RadPrime DNA labeling system (Invitrogen). Vancomycin-resistant isolates were tested for the presence of *vanA* and *vanB* genes using the oligonucleotides and conditions described elsewhere (Dutka-Malen et al., 1995). (Galloway-Pena et al., 2009)

TABLE 1. Primers used in this study for colony hybridizations of putative MSCRAMMs and virulence factors.

Oligonucleotide	Sequence 5′-3′	Length of the amplicon (bp)
Fms1F	ACCAAAGCCAGACGAAATAGAAGAAG	898
Fms1R	ATTGTTTTGGTCAGGTGCATCATAGA	
Fms5F	ATGTTTACCGCAGAAGCAAC	766
Fms5R	ACTTGTATCCGTTGGCTGTT	
Fms9F	GAGGAGACAGCAGCTCAAG	1674
Fms9R	TGTACCTTTGTGTTTATTTGGTA	
Fms10F	GAAACATACATTCCTAAAGACCCTGG	921

Fms10R	ACCATTGTCGCTGATAAATATT	
Fms11F	TGCTAACCAAACGACAGAGGAGAC	832
Fms11R	TTGTGCAAAAGAATAGCCCGTTAC	
Fms13F	GAAGAAGTCGCACAAAAAC	1494
Fms13R	TGACCCGCGTTTCATATTCG	
Fms14F	CTGGCGAAATTACGACTATGCTACTA	924
Fms14R	GTCGCTTTATTATCTCCTTCTTTTAC	
Fms15F	GAGTCTTTAGCAGAACAGCCAGAGCG	1734
Fms15R	TCGATTGTCACTATTCACAAG	
Fms16F	GCCGTATCACCTACACCAG	1206
Fms16R	GATTCCTTTAGGTTGGTTATC	
Fms17F	ATGAAAATGATGGCTTGGCT	828
Fms17R	GGATGATGACCTCGATTCTC	
Fms18F	ATTGATGTGGAAACTGAGGG	1227
Fms18R	TCCTTCTGGAGCCTCTACT	
Fms19F	GTGTGGAAGACGCACAAAGA	350
Fms19R	GGGACTTTATCCCCATCTGC	
Fms20F	CATTGTTTGGGTTAAAAACAG	804
Fms20R	TCAGCTTAGAGCTTCCTCCT	
Fms21F	CTTATTGGAATGTTAGGAATCAT	758
Fms21R	TCAGTAGCAGTCAGCTTTCC	
Esp_{Efm}F	TTGCTAATGCTAGTCCACGACC	945
Esp_{Efm}R	GCGTCAACACTTGCATTGCCGA	
Hyl_{Efm}F	GAGTAGAGGAATATCTTAGC	661
Hyl_{Efm}R	AGGCTCCAATTCTGT	
AcmF	CAGGCAGAGATATCAGCAG	1474
AcmR	ATTCTCATTTGTAACGACTAGC	

Table from (Galloway-Pena et al., 2009)

[2.2.6] Statistical analysis.

The difference between CC17 and non-CC17 isolates, in terms of distributions of the *fms* genes, hyl_{Efm}, esp_{Efm}, ampicillin resistance, and vancomycin resistance were compared using 2-tailed Fisher's exact test. $P < 0.05$ was considered to denote statistical significance. (Galloway-Pena et al., 2009)

[2.2.7] Bacterial isolates and *pbp5* gene sequencing.

Twenty-nine isolates from our previous study (Galloway-Pena et al., 2009) were chosen to represent each major outbreak, ampicillin MIC, and sequence type by MLST; these isolates were collected between 1971 and 1993. The complete *pbp5* gene (GenBank accession no. ZP_00603984) of the 29 isolates was sequenced using the primers found in Table 2. Many of these isolates were known to be in the CC17 genogroup. Twelve additional strains isolated between 1985 and 2006 were studied because they have sequenced genomes or a sequenced *pbp5* region. (Galloway-Pena et al., 2011)

[2.2.8] Other gene comparisons and phylogenetic studies.

Three other genes were studied to determine if the variability found in *pbp5* was comparable to the variability of other genes that encode putative surface proteins within the same genomes and if the isolates with these genes grouped the same with these genes as with *pbp5*. The additional genes analyzed were *pbp2* (GenBank accession no. ZP_03980638), *gls20* (Choudhury et al., 2011)(GenBank accession no. ZP_03981586), and *wlcA* (GenBank accession no. ZP_03980840) (a gene we are currently studying which is related to the *Enterococcus faecalis* WxL genes (Brinster et al., 2007a; Brinster et al., 2007b)). Using the sequence from *E. faecium* strain TX1330, we blasted the accessible genomes of *E. faecium* isolated in the United States available on the Broad Institute or NCBI website for each gene. Each of the genes was aligned using ClustalW and the percent difference between the nucleotide and amino acids analyzed. The nucleotide sequence for each gene was then analyzed using MEGA 4.0.2 software and phylogenetic UPGMA (unweighted-pair group method using average linkages) trees were constructed for each gene (see 2.3.2 Figures 4 and 5). Only strains with sequenced genomes isolated in the United States were used, as this began as a study to compare the sequenced U.S. isolates. In addition, we also did not include any sequences of genes that were ambiguous, poor sequence quality, or incomplete. (Galloway-Pena et al., 2011)

TABLE 2. Primers used in this study to sequence the complete *pbp5* gene of *Enterococcus faecium*.

Primer Name	Primer Sequence (5'-3')	Amplicon Size
Efmpbp5-1outsideF	GGAATGACAAGCAAGAGAAGGAGG	522nts
Efmpbp5-1F	ATGAAAAGAAGTGACAAGCACGGC	493nts
Efmpbp5-1R	CTTACTTTGTCATTTCCTTC	
Efmpbp5-2F	GAAGGAAATGACAAAGTAAG	466nts
Efmpbp5-2R	GTCCCACGAAGATCCTTATCAAAAGCC	
Efmpbp5-3F	GGCTTTTGATAAGGATCTTCGTGGGAC	563nts
Efmpbp5-3R	CCCATTTTCAACGTTTCTTGTGCC	
Efmpbp5-4F	GGCACAAGAAACGTTGAAAATGGG	584nts
Efmpbp5-4R	TTATTGATAATTTTGGTTGAGGTATTG	
Efmpbp5-4outsideR	CGCCACAGTCCTTTTACTGTAC	619nts
Rpbp5_1F	GCAAAGATGAATACCTCATTAGG	357nts
Rpbp5_1R	CAAAGTAATCGGGTTGTACCCAGC	
Rpbp5_2F	CAGAACTTCCAGCTGGAGCTAC	413nts
Rpbp5_2R	GATCATAGCTTGGAGAGCTAGC	
Rpbp5_3F	GCGACAGGTTATGCTCCTGG	423nts
Rpbp5_3R	GAATACATTGCTGCTTGCTGGATAGG	

Table from (Galloway-Pena et al., 2011)

[2.3] CHAPTER 1 RESULTS

[2.3.1] Analysis of Clonality and Antibiotic Resistance among Early Clinical Isolates of *Enterococcus faecium* in the United States.

Research on the emergence of ampicillin and vancomycin resistance and the origins of hospital-associated genogroups are mainly European studies (Leavis et al., 2006a; Lester et al., 2008; Willems et al., 2005). Outbreaks of ampicillin-resistant *E. faecium* were reported as early as the 1980's in the United States (Bush et al., 1989; Coudron et al., 1984; Sapico et al., 1989), yet studies on the population genetics of early ampicillin and vancomycin resistant isolates from the United States have not been done. In this study, I analyzed early United States *E. faecium* for the emergence of putative virulence genes and antibiotic resistance in relation to their clonality in order to better understand the epidemiology of early ampicillin resistant *E. faecium* outbreaks in the United States. A chronological timeline summarizing the key events discussed below can be found in Figure 1. Additionally, illustrations summarizing the comparisons between CC17 isolates and non-CC17 isolates for the presence of ampicillin resistance, vancomycin resistance, and putative virulence genes can be found in Figures 2 and 3.

> Although not a prospective collection, the isolates used in the present study (Table 3) represent isolates causing most of the published *E. faecium* outbreaks occurring up to 1994, as well as early reports of ampicillin-resistant and vancomycin-resistant strains in the United States. In chronological order, the earliest isolates available to us were from the 1970s and were recovered at the Massachusetts General Hospital in Boston during a 22-year study at this institution (Grayson et al., 1991). The isolates recovered in the 1970s were both ampicillin and vancomycin susceptible, and MLST analysis revealed that they did not belong to the CC17 genogroup. Instead they were ST296, ST25, or *purK* allele 8 (NEDH 4901 and NEDH 4586 have an identical pulsotype; data not shown). Colony hybridization analysis showed that these isolates did not contain the putative virulence determinants hyl_{Efm} or esp_{Efm} and harbored only 8 to 9 of the 14 *fms* genes (Table 4). High-level resistance to streptomycin was displayed in 2 of these isolates.

Two published studies reported infections due to ampicillin-resistant *E. faecium* between 1981 and 1987. We obtained 7 isolates from one study performed at Rancho Los Amigos Medical Center in Downey, California, between January 1981 and September 1987 (Miranda et al., 1991; Sapico et al., 1989). This was one of the first reports of an increased presence of ampicillin resistance (MICs 16- 32 µg/mL). We found high-level resistance to streptomycin in some of the isolates; however, none had high-level resistance to gentamicin, and all were vancomycin susceptible. Six of the isolates were ST92, ST476 or contained *purK* allele 2 (RLA-5 and RLA-7 had an identical pulsotype, whereas RLA-1, RLA-2, and RLA-3 had an identical pulsotype; data not shown and (Miranda et al., 1991)), which indicated that they were not related to the CC17 genogroup. The hyl_{Efm} and esp_{Efm} genes were not present, and the *fms* gene profile showed these isolates were missing the predicted pilus operon *fms11-19-16*, putative accessory pilin *fms20*, and putative adhesin *fms18*. One isolate was ST280 (within the CC17 genogroup) and contained all 14 of the *fms* genes; the exact year (between 1981 and 1987) this organism was isolated is unknown.

We also found that CC17 genogroup isolates from the second study performed during this time frame were responsible for an *E. faecium* outbreak in a neonatal intensive care unit at the Medical College of Virginia in Richmond in 1982 (Coudron et al., 1984; Miranda et al., 1991). This outbreak was the first for which bacteremia and meningitis due to multi-resistant *E. faecium* were described. Seven isolates from this outbreak were studied, and we found that all either belonged to ST17, as determined by MLST, or harbored the *purK* allele 1 (MCV161 and 130 had an identical pulsotype, whereas MCV264, MCV211, and MCV255 had an identical pulsotype; data not shown and (Miranda et al., 1991)). This outbreak was different from the California outbreak in that ampicillin MICs were slightly higher, ranging from 32 µg/mL to 64 µg/mL. This is the first time that *E. faecium* with high-level resistance to gentamicin was observed in our study and, to our knowledge, in the United States. These early isolates contained all of the *fms* genes; only one isolate from this outbreak had the esp_{Efm} gene, and none had the hyl_{Efm} gene.

We examined ten isolates recovered in 4 different states between 1986 and 1988. Two isolates were from 2 different hospitals in Boston, Massachushetts (Eliopoulos et al., 1988; Grayson et al., 1991), 1 from a nosocomial outbreak of *E. faecium* at Miriam Hospital in Providence, Rhode Island (Boyce et al., 1992), 3 from Charlotte Memorial Hospital in Charlotte, North Carolina (Stratton et al., 1987), and 4 from the University of Wisconsin at Madison clinics (Spiegel, 1988). One isolate from the Children's Hospital Medical Center in Boston, Massachushetts, had been obtained because of its reported high-level resistance to gentamicin (Eliopoulos et al., 1988). This isolate was found to be ST112 and therefore did not belong to CC17; it was isolated in 1986, had an intermediate ampicillin MIC of 8 µg/mL, and lacked the ebp_{fm} operon *fms1-5-9* but had the esp_{Efm} gene. Another isolate from Boston, recovered in 1988 at New England Deaconess Hospital, had a new ST (ST473) with *purK* allele 8, was ampicillin resistant (MIC 32 µg/mL), and lacked the putative adhesins *scm* (*fms 10*), *fms15* and *fms18*. Four of the 10 isolates recovered during this period were within the

CC17 genogroup, specifically ST18, ST19, or *purK* allele 1 (M-1634 and W-27358 had an identical pulsotype; data not shown). The others were ST10, ST25 or a new ST (ST474). Four isolates from locations other than Boston were ampicillin resistant (3 were in the CC17 genogroup and 1 was not) with MICs of 16 µg/mL to 64 µg/mL. The majority (7) of the 10 isolates showed high-level resistance to streptomycin, but only 1 has high-level resistance to gentamicin. All 4 CC17 genogroup isolates contained all *fms* genes, 3 contained the esp_{Efm} gene, and 2 contained hyl_{Efm}. The 2 isolates that contained the hyl_{Efm} gene were isolated in 1986; this is the first appearance of the hyl_{Efm} gene in the present study.

In addition to the 10 isolates described above, 6 isolates from the Medical College of Pennsylvania in Philadelphia, all of which were isolated in 1988, were also studied (Bush et al., 1989). These isolates were from a surveillance study that showed an increase in penicillin resistance and an inability to treat some of these isolates with combination gentamicin/penicillin therapy. We found that 6 isolates were within the CC17 genogroup with ST16, ST17, or *purK* allele 1 (FA295 and FA287 had an identical pulsotype while FA232 and FA191 had a similar pulsotype; data not shown and (Miranda et al., 1991)) and had all 14 *fms* genes. They were all ampicillin resistant and vancomycin susceptible, with ampicillin MICs of 32 -128 µg/mL. High-level resistance to streptomycin was present in 5 of the isolates, whereas high-level resistance to gentamicin was found in 3 of the isolates during this time period. Five of the isolates contained the esp_{Efm} gene, but none contained the hyl_{Efm} gene.

We also had available isolates from 2 published outbreaks occurring during 1990-1991 (Grayson et al., 1991; Handwerger et al., 1993): 4 isolates were obtained from Boston (2 from New England Deaconess Hospital and 2 from the 22-year review study at Massachusetts General Hospital) (Grayson et al., 1991) and 8 isolates were obtained from a nosocomial outbreak of multiresistant *E. faecium* at Beth Israel Hospital in New York (Handwerger et al., 1993). This was the first time we documented vancomycin-resistant isolates in the present study. Isolates from both outbreaks were all representatives of the CC17 genogroup, (specifically ST16-ST18) or harbored the *purK* allele 1 (SH-4, 5 and 10 had identical pulsotypes, whereas SH-6, SH-9, SH-11, SH-13 and SH-14 had similar pulsotypes; data not shown and (Handwerger et al., 1993)). The majority of isolates were ampicillin resistant, with their MICs (range, 32-256 µg/mL) even higher than those for the isolates recovered from outbreaks occurring in the late 1980s. The isolates from Boston were vancomycin susceptible, whereas the isolates from New York exhibited vancomycin resistance conferred by the *vanA* gene with MICs of 256-512 µg/mL. Both outbreaks had isolates with high-level resistance to both gentamicin and streptomycin. Interestingly, all isolates in both outbreaks contained hyl_{Efm} and esp_{Efm}, highlighting a marked increase in the presence of these 2 genes. Furthermore, most contained all 14 *fms* genes, except that none of the isolates from New York had the putative accessory pilus encoding gene *fms20*.

Last, we also studied 8 isolates recovered in Houston, Texas, at 2 different hospitals between 1992 and 1994 (Coque et al., 1996). Five isolates were within the hospital-associated CC17 genogroup, with ST20, ST17, and *purK* allele 1

represented (VREH-1 and 4 had similar pulsed-field types; data not shown). These isolates exhibited ampicillin resistance with MICs of 32-128 µg/mL, and 2 exhibited *vanA*-type vancomycin resistance (MIC, 512µg/ml). The majority of isolates (5 of 8) exhibited high-level resistance to streptomycin, but only 1 isolate had high-level resistance to gentamicin. The Texas CC17 genogroup isolates had the majority of *fms* genes, except for *fms20* and *fms18.* The hyl_{efm} and esp_{efm} genes were detected in only 1 of these isolates. We also identified a new CC17 genogroup-related ST (ST475) in this group of isolates represented by isolate TX1399, a single locus variant of ST17 in the *ddl* allele. This isolate was susceptible to both ampicillin and vancomycin but had the same *fms*, hyl_{Efm}, and esp_{Efm} gene profile as the other CC17 isolates from Texas. Two of the strains that were isolated were not within the CC17 genogroup (ST71 and *purK* allele 8). Both were ampicillin and vancomycin susceptible, and 1 showed high-level resistance to streptomycin; these 2 isolates lacked the hyl_{Efm} and esp_{Efm} genes and contained only 7 or 8 of the 14 *fms* genes. All of the important events that occurred over time in the isolates in this study are summarized in Figure 1. (Galloway-Pena et al., 2009)

TABLE 3. Relevant background and characteristics of the early clinical United States *E. faecium* isolates included in this study.

Isolate Name(s)	Location	Year	Antibiotic resistance(s)[a]	ST[b,c,d,e]	*purK*	References
NEDH 4901/TX2050	Boston, MA	1971	AMP^{S}, VAN^{S}, STR^{HLR}	296	8	(Grayson et al., 1991; Miranda et al., 1991)
NEDH 4586/TX2051	Boston, MA	1971	AMP^{S}, VAN^{S}	(296)	8	(Grayson et al., 1991; Miranda et al., 1991)
NEDH 758/TX2052	Boston, MA	1973	AMP^{S}, VAN^{S}, STR^{HLR}	25	6	(Grayson et al., 1991; Miranda et al., 1991)
RLA-1/TX2058	Downey, CA	1981-87	AMP^{R}, VAN^{S}, STR^{HLR}	476[e]	2	(Miranda et al., 1991; Sapico et al., 1989)
RLA-2/TX2059	Downey, CA	1981-87	AMP^{R}, VAN^{S}, STR^{HLR}	(476)[e]	2	(Miranda et al., 1991; Sapico et al., 1989)
RLA-3/TX2060	Downey, CA	1981-87	AMP^{R}, VAN^{S}, STR^{HLR}	(476)[e]	2	(Miranda et al., 1991; Sapico et al., 1989)
RLA-4/TX2061	Downey, CA	1981-87	AMP^{R}, VAN^{S}, STR^{HLR}	92	2	(Miranda et

						al., 1991; Sapico et al., 1989)
RLA-5/TX2062	Downey, CA	1981-87	AMP^R, VAN^S	92	2	(Miranda et al., 1991; Sapico et al., 1989)
RLA-7/TX2064	Downey, CA	1981-87	AMP^R, VAN^S	(92)	2	(Miranda et al., 1991; Sapico et al., 1989)
RLA-6/TX2063	Downey, CA	1981-87	AMP^R, VAN^S, STR^{HLR}	280[c]	1	(Miranda et al., 1991; Sapico et al., 1989)
MCV130/TX2029	Richmond, VA	1982	AMP^R, VAN^S	17[c]	1	(Coudron et al., 1984; Miranda et al., 1991)
MCV161/TX2031	Richmond, VA	1982	AMP^R, VAN^S	(17)[c]	1	(Coudron et al., 1984; Miranda et al., 1991)
MCV211/TX2033	Richmond, VA	1982	AMP^R, VAN^S, STR^{HLR}	17[c]	1	(Coudron et al., 1984; Miranda et al., 1991)
MCV255/TX2035	Richmond, VA	1982	AMP^R, VAN^S	(17)[c]	1	(Coudron et al., 1984; Miranda et al., 1991)

MCV/264/TX2036	Richmond, VA	1982	AMP^{R}, VAN^{S}, STR^{HLR}	(17)[c]	1	(Coudron et al., 1984; Miranda et al., 1991)
MCV266/TX2034	Richmond, VA	1982	AMP^{R}, VAN^{S}, STR^{HLR}	17[c]	1	(Coudron et al., 1984; Miranda et al., 1991)
MCV268/TX2038	Richmond, VA	1982	AMP^{R},VAN^{S}, GEN^{HLR}, STR^{HLR}	17[c]	1	(Coudron et al., 1984; Miranda et al., 1991)
NEDH-87-P1/TX2053	Boston, MA	1986	AMP^{S}, VAN^{S}, GEN^{HLR}, STR^{HLR}	112	18	(Eliopoulos et al., 1988; Miranda et al., 1991)
W-5586/TX2008	Charlotte, NC	1986	AMP^{R}, VAN^{S}, STR^{HLR}	10	6	(Stratton et al., 1987)
W-27358/TX2016	Charlotte, NC	1986	AMP^{R}, VAN^{S}	18[c]	1	(Stratton et al., 1987)
M-1634/TX2000	Charlotte, NC	1986	AMP^{R}, VAN^{S}, STR^{HLR}	(18)[c]	1	(Stratton et al., 1987)
UWHC-9802/TX2041	Madison, WI	1987-88	AMP^{S}, VAN^{S}, STR^{HLR}	25	6	(Miranda et al., 1991; Spiegel, 1988)
UWHC-4827/TX2042	Madison, WI	1987-88	AMP^{S}, VAN^{S}, STR^{HLR}	19[c]	1	(Miranda et al., 1991; Spiegel, 1988)

UWHC-112/TX2043	Madison, WI	1987-88	AMP^{S}, VAN^{S}, STR^{HLR}	25	6	(Miranda et al., 1991; Spiegel, 1988)
UWHC-2145/TX2046	Madison, WI	1987-88	AMP^{S}, VAN^{S}	474[e]	4	(Miranda et al., 1991; Spiegel, 1988)
NEDH SF-76/TX2054	Boston, MA	1988	AMP^{R}, VAN^{S}	473[e]	8	(Grayson et al., 1991; Miranda et al., 1991)
875-D/TX2048	Providence, RI	1988	AMP^{R}, VAN^{S}, STR^{HLR}	18[c]	1	(Boyce et al., 1992)
FA287/TX2024	Philadelphia, PA	1988	AMP^{R}, VAN^{S}, GEN^{HLR}, STR^{HLR}	16[c]	1	(Bush et al., 1989; Miranda et al., 1991)
FA295/TX2023	Philadelphia, PA	1988	AMP^{R}, VAN^{S}, GEN^{HLR}, STR^{HLR}	(16)[c]	1	(Bush et al., 1989; Miranda et al., 1991)
FA280/TX2025	Philadelphia, PA	1988	AMP^{R}, VAN^{S}, STR^{HLR}	16[c]	1	(Bush et al., 1989; Miranda et al., 1991)
FA191/TX2027	Philadelphia, PA	1988	AMP^{R}, VAN^{S}, STR^{HLR}	17[c]	1	(Bush et al., 1989; Miranda et al., 1991)

FA232/TX2022	Philadelphia, PA	1988	AMP^{R}, VAN^{S}, GEN^{HLR}, STR^{HLR}	$(17)^{c}$	1	(Bush et al., 1989; Miranda et al., 1991)
FA185/TX2026	Philadelphia, PA	1988	AMP^{R}, VAN^{S}	17^{c}	1	(Bush et al., 1989; Miranda et al., 1991)
A358/TX2067	Boston, MA	1990	AMP^{R}, VAN^{S}, GEN^{HLR}, STR^{HLR}	17^{c}	1	(Singh et al., 1996)
A490/TX2069	Boston, MA	1990	AMP^{R}, VAN^{S}, GEN^{HLR}, STR^{HLR}	17^{c}	1	This study
A360/TX2068	Boston, MA	1990	AMP^{R}, VAN^{S}, GEN^{HLR}, STR^{HLR}	17^{c}	1	This study
A491/TX2070	Boston, MA	1990	AMP^{R}, VAN^{S}, GEN^{HLR}, STR^{HLR}	17^{c}	1	This study
SH-4/TX2429	New York, NY	1990-91	AMP^{R}, VAN^{S}, GEN^{HLR}, STR^{HLR}	18^{c}	1	(Handwerger et al., 1993)
SH-5/TX2430	New York, NY	1990-91	AMP^{R}, VAN^{R}	$(18)^{c}$	1	(Handwerger et al., 1993)
SH-10/TX2435	New York, NY	1990-91	AMP^{R}, VAN^{R}, GEN^{HLR}, STR^{HLR}	$(18)^{c}$	1	(Handwerger et al., 1993)
SH-11/TX2436	New York, NY	1990-91	AMP^{R}, VAN^{R}, GEN^{HLR}, STR^{HLR}	16^{c}	1	(Handwerger et al., 1993)
SH-6/TX2431	New York, NY	1990-91	AMP^{S}, VAN^{R}, GEN^{HLR}, STR^{HLR}	$(16)^{c}$	1	(Handwerger et al., 1993)
SH-9/TX2434	New York, NY	1990-91	AMP^{R}, VAN^{R}, GEN^{HLR}, STR^{HLR}	$(16)^{c}$	1	(Handwerger et al., 1993)
SH-13/TX2438	New York, NY	1990-91	AMP^{R}, VAN^{R}, STR^{HLR}	$(16)^{c}$	1	(Handwerger

						et al., 1993)
SH-14/TX2439	New York, NY	1990-91	AMP^{R}, VAN^{R}, GEN^{HLR}, STR^{HLR}	(16)[c]	1	(Handwerger et al., 1993)
VREH-1/TX2416	Houston,TX	1992	AMP^{R}, VAN^{R}, GEN^{HLR}, STR^{HLR}	20[c]	1	(Korten et al., 1994)
VREH-4/TX2419	Houston, TX	1992	AMP^{R}, VAN^{R}	(20)[c]	1	This study
VREH-5/TX2420	Houston, TX	1992	AMP^{R}, VAN^{S}, STR^{HLR}	17[c]	1	This study
SEH1a/TX1361	Houston, TX	1993-94	AMP^{S}, VAN^{S}, STR^{HLR}	71	2	(Coque et al., 1996)
SEH7/TX1371	Houston, TX	1993-94	AMP^{R}, VAN^{S}, STR^{HLR}	17[c]	1	(Coque et al., 1996)
SEH33/TX1399	Houston, TX	1993-94	AMP^{S}, VAN^{S}	475[c,e]	1	(Coque et al., 1996)
SEH41/TX1405	Houston, TX	1993-94	AMP^{R}, VAN^{S}, STR^{HLR}	17[c]	1	(Coque et al., 1996)
SEH36/TX1401	Houston, TX	1993-94	AMP^{S}, VAN^{S}	ND[d]	8	(Coque et al., 1996)

[a]Superscript "R" denotes resistance, "S" denotes susceptible, and "HLR" denotes high-level resistance, [b]ST numbers in parenthesis denote the inferred sequence type based on having the same *purK* allele and an identical pulsotype to a typed isolate, [c]ST numbers that are considered to be within the CC17 genogroup, [d]ND denotes not done, [e]New STs designated in this study.

Adapted from (Galloway-Pena et al., 2009)

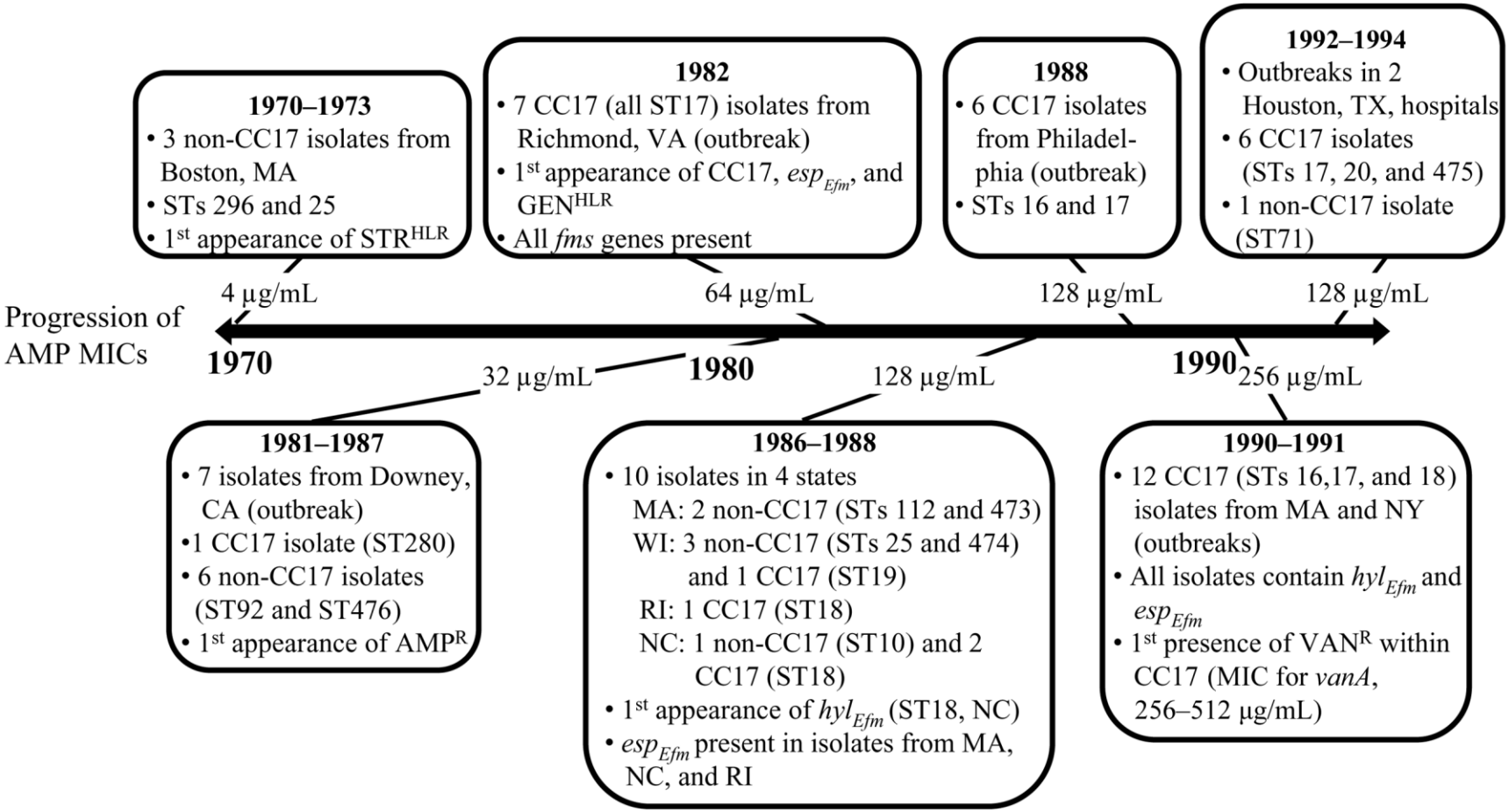


FIGURE 1. **Key events that occurred in the early clinical United States isolates of this study.** The MICs listed correspond to the highest ampicillin MIC from each outbreak. The following abbreviations are used in this figure: high-level resistance to streptomycin

(STR^{HLR}), high-level resistance to gentamicin (GEN^{HLR}), ampicillin resistance (AMP^{R}) and vancomycin resistance (VAN^{R}). Figure from (Galloway-Pena et al., 2009).

TABLE 4. Results for early clinical United States isolate colony hybridizations.

Isolate name/alternate designation	*esp*$_{Efm}$	*hyl*$_{Efm}$	*fms 1*	*fms 5*	*fms 9*	*fms 11*	*fms 19*	*fms 16*	*fms 14*	*fms 17*	*fms 13*	*fms 21*	*fms 20*	*Scm*[a]	*fms 15*	*fms 18*	*acm*[b]
NEDH 4901/TX2050	–	–	+	+	+	+	+	+	+	–	–	+	+	–	–	–	+
NEDH 4586/TX2051	–	–	+	+	+	+	+	+	+	–	–	+	+	–	–	–	+
NEDH 758/TX2052	–	–	–	–	–	–	–	–	+	+	+	+	+	+	+	+	+
RLA-1/TX2058	–	–	+	+	+	–	–	–	–	–	+	+	–	+	+	–	+
RLA-2/TX2059	–	–	+	+	+	–	–	–	–	–	+	+	–	+	+	–	+
RLA-3/TX2060	–	–	+	+	+	–	–	–	–	–	+	+	–	+	+	–	+
RLA-4/TX2061	–	–	+	+	+	–	–	–	+	+	+	+	–	+	+	–	+
RLA-5/TX2062	–	–	+	+	+	–	–	–	–	–	+	+	–	+	+	–	+
RLA-7/TX2064	–	–	+	+	+	–	–	–	–	–	+	–	–	+	+	–	+
RLA-6/TX2063	–	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
MCV130/TX2029	–	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
MCV161/TX2031	–	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
MCV211/TX2033	–	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
MCV255/TX2035	–	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
MCV/264/TX2036	–	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
MCV266/TX2034	–	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
MCV268/TX2038	+	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
NEDH-87-P1/TX2053	+	–	–	–	–	+	+	+	–	+	+	+	+	+	+	+	+
W-5586/TX2008	–	–	–	–	–	–	–	–	+	+	+	+	+	+	+	–	+
W-27358/TX2016	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
M-1634/TX2000	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
UWHC-9802/TX2041	–	–	–	–	–	–	–	–	+	+	+	–	–	+	+	–	+
UWHC-4827/TX2042	–	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
UWHC-112/TX2043	–	–	+	+	+	+	+	+	+	+	+	+	+	+	+	–	+
UWHC-2145/TX2046	–	–	–	–	–	–	–	–	+	+	–	+	+	+	–	–	+
NEDH SF-76/TX2054	–	–	+	+	+	+	+	+	+	+	+	+	+	–	–	–	+
875-D/TX2048	+	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
FA287/TX2024	+	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
FA295/TX2023	+	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
FA280/TX2025	+	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
FA191/TX2027	+	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
FA232/TX2022	+	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
FA185/TX2026	–	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
A358/TX2067	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
A490/TX2069	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
A360/TX2068	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
A491/TX2070	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

SH-4/TX2429	+	+	+	+	+	+	+	+	+	+	+	+	–	+	+	+	+
SH-5/TX2430	+	+	+	+	+	+	+	+	+	+	+	+	–	+	+	+	+
SH-10/TX2435	+	+	+	+	+	+	+	+	+	+	+	+	–	+	+	+	+
SH-11/TX2436	+	+	+	+	+	+	+	+	+	+	+	+	–	+	+	+	+
SH-6/TX2431	+	+	+	+	+	+	+	+	+	+	+	+	–	+	+	+	+
SH-9/TX2434	+	+	+	+	+	+	+	+	+	+	+	+	–	+	+	+	+
SH-13/TX2438	+	+	+	+	+	+	+	+	+	+	+	+	–	+	+	+	+
SH-14/TX2439	+	+	+	+	+	+	+	+	+	+	+	+	–	+	+	+	+
VREH-1/TX2416	+	+	+	+	+	+	+	+	+	+	+	+	–	+	+	+	+
VREH-4/TX2419	–	–	+	+	+	+	+	+	+	+	+	+	–	+	+	+	+
VREH-5/TX2420	–	–	+	+	+	+	+	+	+	+	+	+	+	+	+	–	+
SEH1a/TX1361	–	–	+	+	+	–	–	–	+	+	+	–	–	+	+	–	+
SEH7/TX1371	–	–	+	+	+	+	+	+	+	+	+	+	+	+	+	–	+
SEH33/TX1399	–	–	+	+	+	+	+	+	+	+	+	+	+	+	+	–	+
SEH41/TX1405	–	–	+	+	+	+	+	+	+	+	+	+	+	+	+	–	+
SEH36/TX1401	–	–	–	–	–	+	+	+	+	+	+	+	–	–	–	–	+

[a]The *scm* gene is *fms10*, [b]The *acm* gene is *fms8*

Table from (Galloway-Pena et al., 2009)

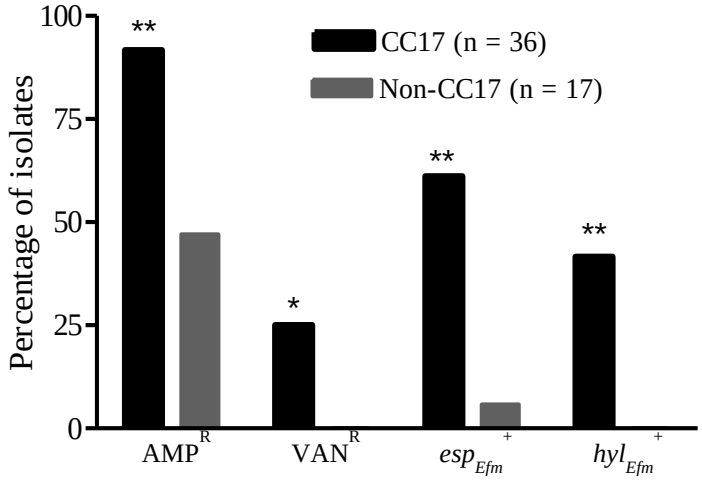


FIGURE 2. A comparison between early United States CC17 isolates and non-CC17 isolates for the presence of ampicillin resistance, vancomycin resistance, the *esp*$_{Efm}$ gene, and the *hyl*$_{Efm}$ gene. * $P < 0.05$ against non-CC17 isolates in our study. ** $P < 0.005$ against non-CC17 isolates in our study. *** $P < 0.0005$ against non-CC17 isolates in our study. Figure from (Galloway-Pena et al., 2009).

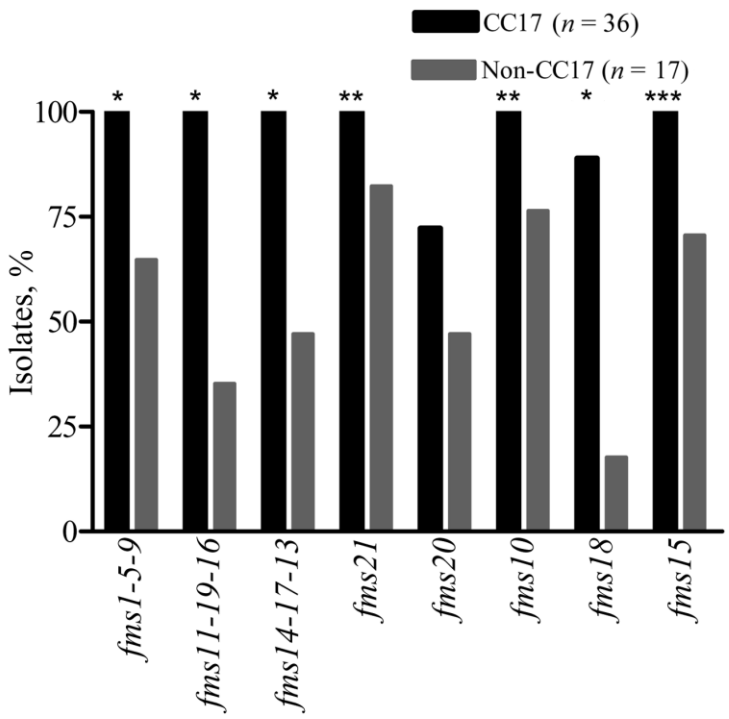


FIGURE 3. A comparison between CC17 and non-CC17 isolates for the presence of the putative adhesins and pili (*fms)* genes in early United States *E. faecium* isolates. *fms1-5-9* is the ebp_{fm} operon while *fms11-19-16* and *fms14-17-13* are putative pilus operons. *fms21* has been shown to encode pili while *fms20* is considered its accessory protein. *fms10*, also known as *scm*, is a collagen binding adhesin while *fms18* and *fms15* are thought to encode putative adhesins.* $P < 0.05$ against non-CC17 isolates in our study. ** $P < 0.005$ against non-CC17 isolates in our study. *** $P < 0.0005$ against non-CC17 isolates in our study. Figure from (Galloway-Pena et al., 2009)

[2.3.2] Molecular Epidemiology of Vancomycin-Resistant *Enterococcus faecium* in South American Hospitals.

VREfm infections have been reported in South America since 1998 (Dalla Costa et al., 1998; Marin et al., 1998). The overall prevalence of VREfm in South America has been found to be much lower (9.7%) than that in the United States (Arias et al., 2003; Moet et al., 2007). The emergence of CC17-related *E. faecium* in South America (Khan et al., 2009; Lopez et al., 2009; Willems et al., 2005) introduced the need for a prospective study addressing the limited data regarding the population genetics of enterococci circulating in the northern region of South America. Therefore, in collaboration, I helped to characterize 35 VREfm isolated in the Andean region of South America at the molecular level.

Of the 760 enterococci isolates collected, 35 *E. faecium* isolates were vancomycin resistant via the *vanA* gene cluster. Eleven different PFGE banding patterns were found with a single PFGE banding pattern prevalent in each country (Table 5) (although some isolates from different countries had related PFGE patterns). MLST analysis of isolates representing each PFGE type revealed that the most frequent ST was ST412. Other STs seen in northern South America included ST17, ST18, ST125, ST203, ST280, ST282, and ST494, all of which are related in the CC17 genogroup.

As done with the United States isolates, the thirty-five South American VREfm isolates were investigated for the presence of various putative virulence genes frequently found in CC17 isolates using DNA probes found in Table 1 for hyl_{Efm}, esp_{Efm}, and the 14 *fms* genes as discussed in section [2.2.5]. The results of colony hybridizations are shown in Figure 4. The four putative MSCRAMM clusters were often present with the *fms5-6-9* (*pilB*) and *fms11-19-16* putative pilus clusters present in all South American VREfm isolates (Figure 4). The *scm*

gene, encoding a collagen adhesin, was present in more than 90% of the isolates. All other *fms* genes were highly represented at rates ranging from 67% to 97% of the VREfm isolates. Additionally, the esp_{Efm} and hyl_{Efm} genes were detected in 69% and 23% of isolates, respectively.

TABLE 5. Molecular typing of the vancomycin resistant *E. faecium* isolates collected from the Andean region of South America.

Country of Origin	PFGE Type	MLST
Colombia	**A**	**412**
Colombia	**A**	
Colombia	**C**	**18**
Colombia	**C**	
Colombia	**C**	
Colombia	**C**	
Colombia	**C**	
Colombia	**C**	
Colombia	**C**	
Colombia	**E**	**125**
Ecuador	**D**	**17**
Ecuador	**D**	
Ecuador	**D**	
Ecuador	**D**	
Ecuador	**I**	**203**
Peru	**A**	**412**
Peru	**B**	**412**
Peru	**B**	
Peru	**B**	
Peru	**B**	
Peru	**B**	
Peru	**B**	
Peru	**E**	**125**
Peru	**F**	**280**
Peru	**F**	
Peru	**G**	**280**
Peru	**H**	**282**
Peru	**J**	**494**
Peru	**K**	**18**
Venezuela	**A**	**412**

Venezuela	**A**
Venezuela	**A**
Venezuela	**A**
Venezuela	**A**
Venezuela	**A**

Adapted from (Panesso et al., 2010)

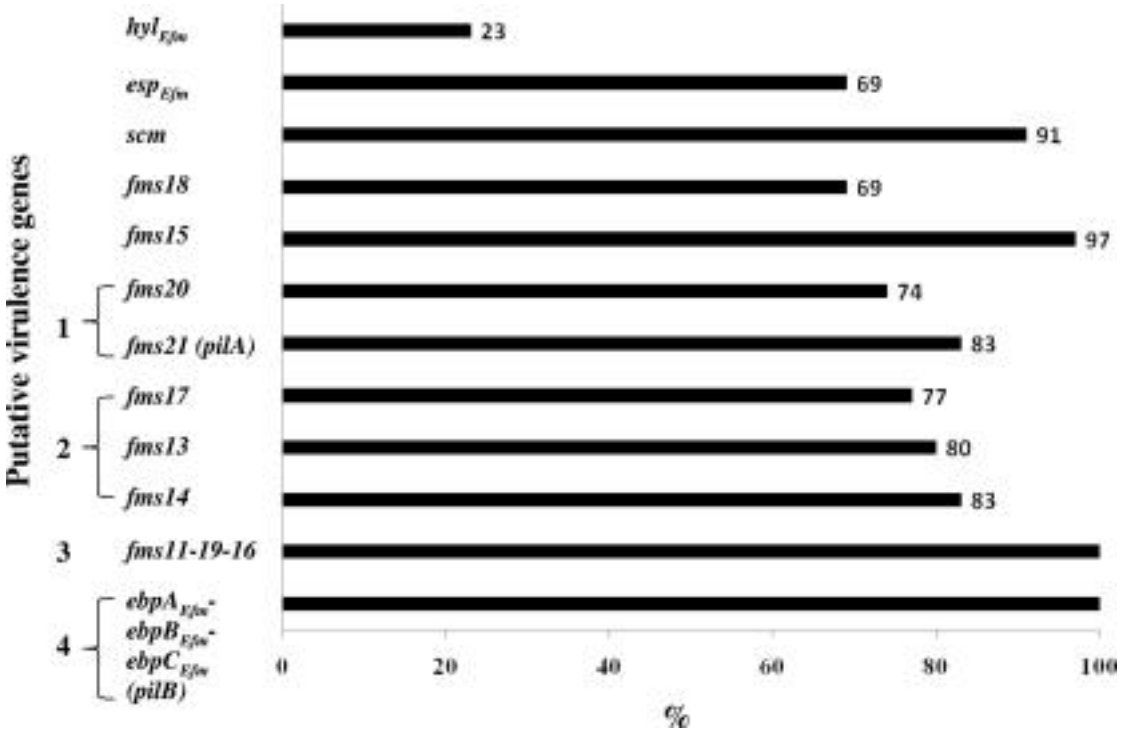


FIGURE 4. **Frequency of putative virulence genes in 35 VR *E. faecium* isolates from the Andean region of South America.** The numbers on the left-hand side indicate the four clusters of genes encoding characterized putative proteins of the enterococcal pili. Figure from (Panesso et al., 2010)

[2.3.3] Analysis of PBP5 of Early United States *Enterococcus faecium* Isolates: Sequence Variation Alone Does Not Explain Increasing Ampicillin Resistance Over Time.

In the first study (section 2.3.1), it was found that ampicillin MICs of hospital-associated *E. faecium* isolates from the United States have increased steadily over the past thirty years. Isolates were ST17, or a related ST. Thus, this study (Galloway-Pena et al., 2011) tried to determine if specific amino acid variations could be correlated with increasing MICs using strains from similar lineages (an attempt to reduce background differences).

> Up to 98 single nucleotide polymorphisms (Table 6), up to 73 silent polymorphisms (Table 7) and up to 20 amino acid variations (Table 8) were found when comparing isolates with an ampicillin MIC < 4µg/ml and those with an MIC of ampicillin >4µg/ml, resulting in an approximately 5% nucleotide difference. All strains with MICs < 4µg/ml (*pbp5*-S group) had an almost identical amino acid profiles, while all strains with MICs >4µg/ml (*pbp5*-R group) had amino acid profiles which were different from the first group and which were almost identical to each other and to the transferable low affinity PBP5 found in the *E. faecium* strain C68 (~99-100% sequence identity) (Carias et al., 1998; Rice et al., 2001). The consensus sequence (Table 8) of the amino acids that differed between the two groups was derived for the strains with an MIC< 4µg/ml and those with an MIC >4µg/ml, and is consistent with the sequences previously reported (Ligozzi et al., 1996; Rice et al., 2001; Rybkine et al., 1998).
>
> By analyzing the silent polymorphisms sites found when comparing *pbp5*-S and *pbp5*-R strains, we found that *pbp5*-R strains use the same codon at all silent polymorphism sites whereas the *pbp5*-S and hybrid strains may use two different codons for an amino acid at a particular position (18 out of 73 silent polymorphisms) (Table 6 and 7). (Galloway-Pena et al., 2011)

TABLE 6. Single nucleotide polymorphisms at each nucleotide position of the consensus sequence of *pbp5* in the 29 United States strains analyzed. Adapted from (Galloway-Pena et al., 2011). See next 3 pages. Variation within the same group (*pbp5*-S or *pbp5*-R) is highlighted in gray).

Strain	63	71	78	79	101	108	144	168	174	197	202	207	216	255	264	282	298	306	330	342	345	429	430	450	453	495	504	514	525	529	543	561	585	594	603
1231501	T	T	A	A	G	C	G	A	A	G	G	G	A	A	A	C	G	A	T	T	C	A	A	G	T	T	G	A	C	T	G	C	T	C	G
Com15	T	T	A	A	G	C	G	A	A	G	G	G	A	A	A	C	G	A	T	T	C	A	A	G	T	T	G	A	C	T	G	C	C	C	G
Com12	T	T	A	A	G	C	G	A	A	G	G	G	A	A	A	C	G	A	T	T	C	A	A	G	T	T	G	A	C	T	G	C	C	C	G
TX1401	T	T	A	A	G	C	G	A	A	G	G	G	A	A	A	C	G	A	T	T	C	A	A	G	T	T	G	A	C	T	G	C	T	C	G
1141733	T	T	A	A	G	C	G	A	A	G	G	G	A	A	A	C	G	A	T	T	C	A	A	G	T	T	G	A	C	T	G	C	C	C	G
TX2050	T	T	A	A	G	C	G	A	A	G	G	G	A	A	A	C	G	A	T	T	C	A	A	G	T	T	G	A	C	T	G	C	T	C	G
TX2042	C	C	G	G	A	T	G	G	A	A	G	A	G	G	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
TX2043	C	C	G	G	A	T	A	G	G	A	G	A	G	G	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
D366	C	C	G	G	A	T	A	G	G	A	G	A	G	G	G	T	C	T	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
TX1399	C	C	G	G	A	T	A	G	G	A	A	A	G	T	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
TX2053	C	C	G	G	A	T	A	G	G	A	A	A	G	T	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
TX0016	C	C	G	G	A	T	A	G	G	A	A	A	G	T	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
TX2058	C	C	G	G	A	T	A	G	G	A	A	A	G	T	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
TX2062	C	C	G	G	A	T	A	G	G	A	A	A	G	T	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
TX2063	C	C	G	G	A	T	A	G	G	A	A	A	G	T	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
TX2008	C	C	G	G	A	T	A	G	G	A	A	A	G	T	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
TX2061	C	C	G	G	A	T	A	G	G	A	A	A	G	T	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
TX2029	C	C	G	G	A	T	A	G	G	A	A	A	G	T	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
TX2067	C	C	G	G	A	T	A	G	G	A	A	A	G	T	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
TX2016	C	C	G	G	A	T	A	G	G	A	A	A	G	T	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
TX1371	C	C	G	G	A	T	A	G	G	A	A	A	G	T	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
TX2033	C	C	G	G	A	T	A	G	G	A	A	A	G	T	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
TX2034	C	C	G	G	A	T	A	G	G	A	A	A	G	T	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
TX2038	C	C	G	G	A	T	A	G	G	A	A	A	G	T	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
TX2024	C	C	G	G	A	T	A	G	G	A	A	A	G	T	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
TX2025	C	C	G	G	A	T	A	G	G	A	A	A	G	T	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
TX2027	C	C	G	G	A	T	A	G	G	A	A	A	G	T	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
TX2068	C	C	G	G	A	T	A	G	G	A	A	A	G	T	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
TX2069	C	C	G	G	A	T	A	G	G	A	A	A	G	T	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
TX1405	C	C	G	G	A	T	A	G	G	A	A	A	G	T	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
1230933	C	C	G	G	A	T	A	G	G	A	A	A	G	T	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
1231408	C	C	G	G	A	T	A	G	G	A	A	A	G	T	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
1231410	C	C	G	G	A	T	A	G	G	A	A	A	G	T	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
TX2436	C	C	G	G	A	T	A	G	G	A	A	A	G	T	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
TX2416	C	C	G	G	A	T	A	G	G	A	A	A	G	T	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
TX2420	C	C	G	G	A	T	A	G	G	A	A	A	G	T	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
TX2022	C	C	G	G	A	T	A	G	G	A	A	A	G	T	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
1231502	C	C	G	G	A	T	A	G	G	A	A	A	G	T	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
TX2070	C	C	G	G	A	T	A	G	G	A	A	A	G	T	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
C68	C	C	G	G	A	T	A	G	G	A	A	A	G	T	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A
H80721	C	C	G	G	A	T	A	G	G	A	A	A	G	T	G	T	C	G	C	C	T	T	C	A	C	C	A	G	T	A	T	A	C	T	A

Strain	609	611	636	645	646	648	681	771	819	864	897	900	903	912	942	951	963	970	981	990	1002	1011	1017	1020	1047	1068	1095	1194	1201	1224	1281	1362	1377	1389	1396
1231501	G	A	C	T	G	A	T	C	A	A	A	A	T	A	T	A	C	A	C	G	G	C	C	A	A	C	G	C	G	A	C	T	G	T	T
Com15	G	A	C	T	G	A	T	C	A	A	A	A	T	A	T	A	C	A	C	G	G	C	C	A	A	C	G	C	G	A	C	T	G	T	T
Com12	G	A	C	T	G	A	T	C	A	A	A	A	T	A	T	A	C	A	C	G	G	C	C	A	A	C	G	C	G	A	C	T	G	T	T
TX1401	G	A	C	T	G	A	T	C	A	A	A	A	T	A	T	A	C	A	C	G	G	C	C	A	A	C	G	C	G	A	C	T	G	T	T
1141733	G	A	C	T	G	A	T	C	A	A	A	A	T	A	T	A	C	A	C	G	G	C	C	A	A	C	G	C	G	A	C	T	G	T	T
TX2050	G	A	C	T	G	A	T	C	A	A	A	A	T	A	T	A	C	A	C	G	G	C	C	A	A	C	G	C	G	A	C	T	G	T	T
TX2042	A	A	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	G	G	T	C	T	G	C
TX2043	A	A	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	T	G	T	C	T	G	C
D366	A	G	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	T	A	T	C	T	G	C
TX1399	A	G	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	G	G	T	C	T	G	C
TX2053	A	G	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	G	G	T	C	T	G	C
TX0016	A	G	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	G	G	T	C	T	G	C
TX2058	A	G	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	G	G	T	C	T	G	C
TX2062	A	G	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	G	G	T	C	T	G	C
TX2063	A	G	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	G	G	T	C	T	G	C
TX2008	A	G	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	G	G	T	C	T	G	C
TX2061	A	G	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	G	G	T	C	T	G	C
TX2029	A	G	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	G	G	T	C	T	G	C
TX2067	A	G	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	G	G	T	C	T	G	C
TX2016	A	G	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	G	G	T	C	T	G	C
TX1371	A	G	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	G	G	T	C	T	G	C
TX2033	A	G	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	G	G	T	C	T	G	C
TX2034	A	G	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	G	G	T	C	T	G	C
TX2038	A	G	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	G	G	T	C	T	G	C
TX2024	A	G	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	G	G	T	C	T	G	C
TX2025	A	G	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	G	G	T	C	T	G	C
TX2027	A	G	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	G	G	T	C	T	G	C
TX2068	A	G	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	G	G	T	C	T	G	C
TX2069	A	G	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	G	G	T	C	T	G	C
TX1405	A	G	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	G	G	T	C	T	G	C
1230933	A	G	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	G	G	T	C	T	G	C
1231408	A	G	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	G	G	T	C	T	G	C
1231410	A	G	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	G	G	T	C	T	G	C
TX2436	A	G	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	T	A	A	A	T	T	C	G	T	A	T	G	G	T	C	T	G	C
TX2416	A	G	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	G	G	T	C	T	G	C
TX2420	A	G	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	G	G	T	C	T	G	C
TX2022	A	G	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	G	G	T	C	T	G	C
1231502	A	G	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	G	T	T	C	T	G	C
TX2070	A	G	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	G	G	T	C	T	G	C
C68	A	G	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	G	A	T	C	T	G	C
H80721	A	G	G	C	T	C	G	T	G	G	T	C	A	G	A	G	T	G	A	A	A	T	T	C	G	T	A	T	G	A	T	C	T	G	C

Strain	1399	1400	1401	1406	1407	1408	1410	1440	1459	1460	1464	1485	1491	1494	1497	1501	1506	1509	1548	1581	1590	1626	1762	1845	1892	1896	1929	1944	1956	1995	2005	2006	2007	2016
1231501	-	-	-	-	-	-	G	T	A	T	G	G	G	T	C	G	T	G	C	G	T	A	G	A	A	A	C	A	T	T	C	C	C	A
Com15	-	-	-	-	-	-	G	T	A	T	G	G	G	T	C	G	T	G	C	G	T	A	G	A	A	A	C	A	T	T	C	C	C	A
Com12	-	-	-	-	-	-	G	T	A	T	A	G	G	T	C	G	T	G	C	G	T	A	G	A	A	A	C	A	T	T	G	C	C	A
TX1401	-	-	-	-	-	-	G	T	A	T	G	G	G	T	C	G	C	A	T	T	C	G	T	G	T	G	T	T	C	C	T	C	G	G
1141733	-	-	-	-	-	-	G	T	A	T	A	G	G	T	C	G	T	G	C	G	T	A	G	A	A	A	C	A	T	T	C	T	C	A
TX2050	-	-	-	-	-	-	G	T	A	T	G	G	G	T	C	G	T	G	C	G	T	A	G	G	A	A	T	T	C	C	C	C	G	G
TX2042	-	-	-	-	-	-	A	C	A	T	A	T	A	A	T	A	C	A	T	T	C	G	G	G	T	A	T	T	C	C	C	C	C	G
TX2043	-	-	-	-	-	-	A	C	A	T	A	T	A	A	T	A	C	A	T	T	C	G	G	G	A	A	T	T	C	C	C	C	G	G
D366	-	-	-	-	-	-	A	C	A	C	A	T	A	A	T	A	C	A	T	T	C	G	G	G	A	A	T	T	C	C	C	C	G	G
TX1399	-	-	-	-	-	-	A	C	A	C	A	T	A	A	T	A	C	A	T	T	C	G	T	G	T	G	T	T	C	C	T	C	G	G
TX2053	-	-	-	-	-	-	A	C	G	C	A	T	A	A	T	A	C	A	T	T	C	G	G	G	T	G	T	T	C	C	T	C	G	G
TX0016	-	-	-	-	-	-	A	C	A	T	A	T	A	A	T	A	C	A	T	T	C	G	G	G	A	G	T	T	C	C	T	C	G	G
TX2058	-	-	-	A	T	G	A	C	G	C	A	T	A	A	T	A	C	A	T	T	C	G	T	G	T	G	T	T	C	C	T	C	G	G
TX2062	-	-	-	A	T	G	A	C	G	C	A	T	A	A	T	A	C	A	T	T	C	G	T	G	T	G	T	T	C	C	T	C	G	G
TX2063	-	-	-	-	-	-	A	C	G	C	A	T	A	A	T	A	C	A	T	T	C	G	T	G	T	G	T	T	C	C	T	C	G	G
TX2008	A	G	T	-	-	-	A	C	A	C	A	T	A	A	T	A	C	A	T	T	C	G	G	G	T	G	T	T	C	C	T	C	G	G
TX2061	-	-	-	A	T	G	A	C	G	C	A	T	A	A	T	A	C	A	T	T	C	G	T	G	T	G	T	T	C	C	T	C	G	G
TX2029	-	-	-	-	-	-	A	C	A	C	A	T	A	A	T	A	C	A	T	T	C	G	T	G	T	G	T	T	C	C	T	C	G	G
TX2067	-	-	-	-	-	-	A	C	A	C	A	T	A	A	T	A	C	A	T	T	C	G	T	G	T	G	T	T	C	C	T	C	G	G
TX2016	-	-	-	-	-	-	A	C	A	C	A	T	A	A	T	A	C	A	T	T	C	G	T	G	T	G	T	T	C	C	T	C	G	G
TX1371	-	-	-	-	-	-	A	C	A	C	A	T	A	A	T	A	C	A	T	T	C	G	T	G	T	G	T	T	C	C	T	C	G	G
TX2033	-	-	-	-	-	-	A	C	G	C	A	T	A	A	T	A	C	A	T	T	C	G	T	G	T	G	T	T	C	C	T	C	G	G
TX2034	-	-	-	-	-	-	A	C	A	C	A	T	A	A	T	A	C	A	T	T	C	G	T	G	T	G	T	T	C	C	T	C	G	G
TX2038	-	-	-	-	-	-	A	C	A	C	A	T	A	A	T	A	C	A	T	T	C	G	T	G	T	G	T	T	C	C	T	C	G	G
TX2024	-	-	-	-	-	-	A	C	A	C	A	T	A	A	T	A	C	A	T	T	C	G	T	G	T	G	T	T	C	C	T	C	G	G
TX2025	-	-	-	-	-	-	A	C	A	C	A	T	A	A	T	A	C	A	T	T	C	G	T	G	T	G	T	T	C	C	T	C	G	G
TX2027	-	-	-	-	-	-	A	C	A	C	A	T	A	A	T	A	C	A	T	T	C	G	T	G	T	G	T	T	C	C	T	C	G	G
TX2068	-	-	-	-	-	-	A	C	G	C	A	T	A	A	T	A	C	A	T	T	C	G	T	G	T	G	T	T	C	C	T	C	G	G
TX2069	-	-	-	-	-	-	A	C	A	C	A	T	A	A	T	A	C	A	T	T	C	G	T	G	T	G	T	T	C	C	T	C	G	G
TX1405	-	-	-	-	-	-	A	C	A	C	A	T	A	A	T	A	C	A	T	T	C	G	T	G	T	G	T	T	C	C	C	T	C	G
1230933	-	-	-	-	-	-	A	C	G	C	A	T	A	A	T	A	C	A	T	T	C	G	T	G	T	G	T	T	C	C	T	C	G	G
1231408	A	G	T	-	-	-	A	C	G	C	A	T	A	A	T	A	C	A	T	T	C	G	G	G	T	G	T	T	C	C	T	C	G	G
1231410	A	G	T	-	-	-	A	C	G	C	A	T	A	A	T	A	C	A	T	T	C	G	G	G	T	G	T	T	C	C	T	C	G	G
TX2436	-	-	-	-	-	-	A	C	G	C	A	T	A	A	T	A	C	A	T	T	C	G	T	G	T	G	T	T	C	C	T	C	G	G
TX2416	-	-	-	-	-	-	A	C	A	C	A	T	A	A	T	A	C	A	T	T	C	G	T	G	T	G	T	T	C	C	C	T	C	G
TX2420	-	-	-	-	-	-	A	C	A	C	A	T	A	A	T	A	C	A	T	T	C	G	T	G	T	G	T	T	C	C	T	C	G	G
TX2022	-	-	-	-	-	-	A	C	A	C	A	T	A	A	T	A	C	A	T	T	C	G	T	G	T	G	T	T	C	C	T	C	G	G
1231502	A	G	T	-	-	-	A	C	G	C	A	T	A	A	T	A	C	A	T	T	C	G	G	G	T	G	T	T	C	C	T	C	G	G
TX2070	-	-	-	A	T	G	A	C	G	C	A	T	A	A	T	A	C	A	T	T	C	G	T	G	T	G	T	T	C	C	T	C	G	G
C68	A	G	T	-	-	-	A	C	G	C	A	T	A	A	T	A	C	A	T	T	C	G	G	G	T	G	T	T	C	C	T	C	G	G
H80721	A	G	T	-	-	-	A	C	G	C	A	T	A	A	T	A	C	A	T	T	C	G	G	G	T	A	T	T	C	C	T	C	G	G

TABLE 7. Analysis of silent polymorphisms and the codon usage differences between *pbp5*-S, hybrid, and *pbp5*-R strains.

Position	Amino Acid	Strains	Codon Used	Position	Amino Acid	Strains	Codon Used	Position	Amino Acid	Strains	Codon Used
21	Isoleucine	pbp-S	AUU	26	Alanine	pbp-S	GCA	36	Tyrosine	pbp-S	UAC
		Hybrid	AUC			Hybrid	GCG			Hybrid	UAU
		pbp-R	AUC			pbp-R	GCG			pbp-R	UAU
48	Threonine	pbp-S	ACG	56	Leucine	pbp-S	UUA	58	Lysine	pbp-S	AAA
		Hybrid	ACG/ACA			Hybrid	UUG			Hybrid	AAA/AAG
		pbp-R	ACA			pbp-R	UUG			pbp-R	AAG
69	Serine	pbp-S	UCG	72	Alanine	pbp-S	GCA	88	Glutamine	pbp-S	CAA
		Hybrid	UCA			Hybrid	GCG			Hybrid	CAG
		pbp-R	UCA			pbp-R	GCG			pbp-R	CAG
94	Alanine	pbp-S	GCC	102	Serine	pbp-S	UCA	110	Aspartic A	pbp-S	GAU
		Hybrid	GCU			Hybrid	UCG/UCU			Hybrid	GAC
		pbp-R	GCU			pbp-R	UCG			pbp-R	GAC
114	Tyrosine	pbp-S	UAU	115	Serine	pbp-S	AGC	143	Glycine	pbp-S	GGA
		Hybrid	UAC			Hybrid	AGU			Hybrid	GGU
		pbp-R	UAC			pbp-R	AGU			pbp-R	GGU
150	Glutamine	pbp-S	CAG	151	Proline	pbp-S	CCU	165	Serine	pbp-S	AGU
		Hybrid	CAA			Hybrid	CCC			Hybrid	AGC
		pbp-R	CAA			pbp-R	CCC			pbp-R	AGC
168	Threonine	pbp-S	ACG	175	Asparagine	pbp-S	AAC	181	Glycine	pbp-S	GGG
		Hybrid	ACA			Hybrid	AAU			Hybrid	GGU
		pbp-R	ACA			pbp-R	AAU			pbp-R	GGU
187	Threonine	pbp-S	ACC	195	Valine	pbp-S	GUU/GUC	198	Serine	pbp-S	AGC
		Hybrid	ACA			Hybrid	GUC			Hybrid	AGU
		pbp-R	ACA			pbp-R	GUC			pbp-R	AGU
201	Glycine	pbp-S	GGG	203	Glycine	pbp-S	GGG	212	Alanine	pbp-S	GCC
		Hybrid	GGA			Hybrid	GGA			Hybrid	GCG
		pbp-R	GGA			pbp-R	GGA			pbp-R	GCG
215	Serine	pbp-S	UCU	227	Alanine	pbp-S	GCU	257	Isoleucine	pbp-S	AUC
		Hybrid	UCC			Hybrid	GCG			Hybrid	AUU
		pbp-R	UCC			pbp-R	GCG			pbp-R	AUU
273	Glutamine	pbp-S	CAA	288	Lysine	pbp-S	AAA	299	Glycine	pbp-S	GGA
		Hybrid	CAG			Hybrid	AAG			Hybrid	GGU
		pbp-R	CAG			pbp-R	AAG			pbp-R	GGU
300	Arginine	pbp-S	CGA	301	Serine	pbp-S	UCU	304	Glutamic A	pbp-S	GAA
		Hybrid	CGC			Hybrid	UCA			Hybrid	GAG
		pbp-R	CGC			pbp-R	UCA			pbp-R	GAG
314	Threonine	pbp-S	ACU	317	Glycine	pbp-S	GGA	321	Isoleucine	pbp-S	AUC
		Hybrid	ACA			Hybrid	GGG			Hybrid	AUU
		pbp-R	ACA			pbp-R	GGG			pbp-R	AUU

Position	Amino Acid	Strains	Codon Used	Position	Amino Acid	Strains	Codon Used	Position	Amino Acid	Strains	Codon Used
327	Valine	pbp-S	GUC	330	Lysine	pbp-S	AAG	334	Glutamic A	pbp-S	GAG
		Hybrid	GUA			Hybrid	AAA			Hybrid	GAA
		pbp-R	GUA			pbp-R	AAA			pbp-R	GAA
337	Valine	pbp-S	GUC	339	Asparagine	pbp-S	AAC	340	Glycine	pbp-S	GGA
		Hybrid	GUU			Hybrid	AAU			Hybrid	GGC
		pbp-R	GUU			pbp-R	AAU			pbp-R	GGC
349	Alanine	pbp-S	GCA	356	Phenylalanine	pbp-S	UUC	365	Serine	pbp-S	UCG
		Hybrid	GCG			Hybrid	UUU			Hybrid	UCA
		pbp-R	GCG			pbp-R	UUU			pbp-R	UCA
398	Aspartic A	pbp-S	GAC	408	Glutamine	pbp-S	CAA	427	Isoleucine	pbp-S	AUC
		Hybrid	GAU			Hybrid	CAG/CAA			Hybrid	AUU
		pbp-R	GAU			pbp-R	CAG			pbp-R	AUU
454	Aspartic A	pbp-S	GAU	459	Serine	pbp-S	UCG	463	Threonine	pbp-S	ACU
		Hybrid	GAC			Hybrid	UCU			Hybrid	ACG
		pbp-R	GAC			pbp-R	UCU			pbp-R	ACG
465	Valine	pbp-S	GUU	468	Valine	pbp-S	GUG	478	Isoleucine	pbp-S	AUU
		Hybrid	GUC			Hybrid	GUA			Hybrid	AUC
		pbp-R	GUC			pbp-R	GUA			pbp-R	AUC
486	Alanine	pbp-S	GCA/GCG	493	Glycine	pbp-S	GGG	495	Lysine	pbp-S	AAG
		Hybrid	GCA			Hybrid	GGU			Hybrid	AAA
		pbp-R	GCA			pbp-R	GGU			pbp-R	AAA
497	Phenylalanine	pbp-S	UUC	500	Glycine	pbp-S	GGU/GGC	501	Leucine	pbp-S	UUG/UUA
		Hybrid	UUU			Hybrid	GGC			Hybrid	UUA
		pbp-R	UUU			pbp-R	GGC			pbp-R	UUA
514	Isoleucine	pbp-S	AUC/AUU	528	Asparagine	pbp-S	AAU/AAC	540	Glutamine	pbp-S	CAA/CAG
		Hybrid	AUU			Hybrid	AAC			Hybrid	CAG
		pbp-R	AUU			pbp-R	AAC			pbp-R	CAG
613	Proline	pbp-S	CCA/CCG	630	Lysine	pbp-S	AAA/AAG	641	Asparagine	pbp-S	AAC/AAU
		Hybrid	CCG			Hybrid	AAG			Hybrid	AAU
		pbp-R	CCG			pbp-R	AAG			pbp-R	AAU
646	Glycine Com15-E GAA	pbp-S	GGA/GGU	650	Valine	pbp-S	GUU/GUC	663	Threonine	pbp-S	ACU/ACC
		Hybrid	GGU			Hybrid	GUC			Hybrid	ACC
		pbp-R	GGU			pbp-R	GUC			pbp-R	ACC
670	Leucine	pbp-S	UUA/UUG								
		Hybrid	UUG								
		pbp-R	UUG								

Table from (Galloway-Pena et al., 2011)

TABLE 8. Amino Acid Changes Between Ampicillin Susceptible and Resistant *Enterococcus faecium* Stains from the United States.

			Amino Acid Position																				
			24	27	34	66	68	85	100	144	172	177	204	216	324	466'	485	496	499	525	586	629	667
			Consensus: MIC < 4μg/ml																				
MIC	Strain	MLST	V	S	R	G	A	E	E	K	T	L	D	A	T	–	M	N	A	E	V	E	P
0.5-1 (n=3)[a]	1231501	52																					
	Com15	583																					
	Com12	107																					A
2 (n=1)	TX1401[b]	ND[c]																			L	V	S
	1141733	327																					S
4 (n=4)	TX2050	296																				V	
	TX2042[b]	19[d]	A	G	Q	E			Q	Q	A	I		S	A			K	T	D		V	
	TX2043[b]	25	A	G	Q	E			Q	Q	A	I		S	A			K	T	D			
	D366[b]	25	A	G	Q	E			Q	Q	A	I	G	S	A			K	I	D			
			Consensus: MIC > 4μg/ml																				
			A	G	Q	E	T	D	Q	Q	A	I	G	S	A	–	T	K	T	D	L	V	S
8 (n=2)	TX1399	475[d]																					
	TX2053	112																			V		
16 (n=5)	TX0016	18[d]															M				V	E	
	TX2058	476														D	A						
	TX2062	92														D	A						
	TX2063	280[d]															A						
	TX2008	10														S					V		
32 (n=5)	TX2061	92														D	A						
	TX2029	17[d]																					
	TX2067	17[d]																					
	TX2016	18[d]																					
	TX1371	17[d]																					
64 (n=9)	TX2033	17[d]															A						
	TX2034	17[d]																					
	TX2038	17[d]																					
	TX2024	16[d]																					
	TX2025	16[d]																					
	TX2027	17[d]																					
	TX2068	17[d]															A						
	TX2069	17[d]															A						
	TX1405	17[d]																					L
128 (n=7)	1230933	18[d]															A						
	1231408	582[d]														S	A				V		
	1231410	17[d]														S	A				V		
	TX2436	16[d]															A						
	TX2416	20[d]																					L
	TX2420	17[d]																					
	TX2022	ND																					
256 (n=4)	1231502	203[d]														S	A				V		
	TX2070	17[d]														D	A						
	C68	16[d]														S	A				V		
	H80721	ND														S	A				V		

Table from (Galloway-Pena et al., 2011)

TABLE 8. Continued

[a] the number of isolates with a specific ampicillin MIC is indicated by (n=), [b] indicates a "hybrid" type *pbp5* gene, [c] ND denotes that the multi-locus sequence type was not determined, [d] indicates an ST17, single locus variant of ST17, or double locus variant of ST17, and therefore considered to be in the CC17 genogroup

Interestingly, some strains (e.g. TX2042) with an MIC of 4µg/ml had a "hybrid-like" PBP5 (indicated in Table 8) with some of the amino acids matching the consensus of the more susceptible strains, while most of the protein sequence matched the consensus of the resistant strains. One strain with an MIC of 2µg/ml (TX1401) had the consensus sequence of susceptible strains, except for the last three consensus sequence amino acids 586, 629 and 667, which coincided with the consensus of the resistant strains. TX1401 was the only strain predicted to have derived its sequence via homologous recombination using the maximum chi-squared test with putative recombination sites at bases 1888 (P=0.009) and 1587 (P=0.016) with approximately the first three-fourths of the sequence resembling the *pbp5*-S and the last part resembling *pbp5*-R strains.

Amino acid changes near the active site predicted in previous publications (Jureen et al., 2004b; Klibi et al., 2008; Ligozzi et al., 1996; Poeta et al., 2007; Rice et al., 2004; Rybkine et al., 1998; Sifaoui et al., 2001; Zorzi et al., 1996) that were thought to be the cause of high-level resistance include an additional serine after amino acid position 466 (Ser-466') and Met-485-Ala/Thr. Ser-466' was found in three of the four most highly resistant isolates (MIC of ampicillin, 256µg/ml) and two isolates with an ampicillin MIC of 128µg/ml but also in one low-level resistant strain (e.g. TX2008) with an MIC of 16µg/ml, indicating there is not an absolute association between these changes and high-level resistance. Interestingly, some strains (both low-level and high-level ampicillin resistant) also had an insertion of an aspartic acid residue instead of a serine at position 466' (e.g., TX2058 and TX2070), which is located in the loop that forms the outer edge of the active site. Both residues are polar, so they may have the same effect, although aspartic acid is negatively charged, while serine is neutral. It was theorized by Sauvage et al. that the insertion at residue 466' may displace the Val-465 in the active site, reducing accessibility for penicillin, affecting substrate specificity (Sauvage et al., 2002). Although these variations were only found in resistant isolates, whether there was a serine or aspartic acid at this position did not seem to correlate with specific increases in MICs. However, the presence of a Ser-466' or Asp-466' in combination with Ala-485 was found in all four isolates with an MIC of 256µg/ml and therefore, while it may not be sufficient, it may be required to achieve this high level of ampicillin resistance. It is thought that removal of the bulky methionine at position 485 would allow for conformational freedom near the Ser-480-Asp-Asn active site, resulting in a less efficient acylation process (Sauvage et al., 2002). Although there are a number of amino acid differences between *pbp5*-S and *pbp5*-R strains (as well as silent polymorphisms) in the membrane anchor (1-45), N-terminal extension (46-162), and N-terminal domain (163-345) these are unlikely to be of great significance since the transpeptidase domain is at the C-terminus (346-680), although we cannot rule out this possibility. Of note, however, all of the amino acid differences among the resistant strains do occur in the transpeptidase domain.

Outside of the Ser/Asp-466' and Ala-485, no single amino acid variation, or combination thereof, correlated with specific ampicillin MICs or the progressive increase in MICs seen in United States outbreak isolates with an ampicillin MIC > 4ug/ml. Table 8 groups the isolates by MIC and shows how each isolate's PBP5 differs from the consensus sequence. Interestingly, strain background did not seem to correlate with the differences seen in the *pbp5*

sequence, as multiple variations were found within the same ST, and the same sequence variation was found in multiple STs. However, the more resistant strains are of STs found in the hospital-associated CC17 genogroup, as to be expected (Galloway-Pena et al., 2009) (CC17 genogroup isolates are footnoted in Table 8).

The finding that most *E. faecium* strains with an MIC >4µg/ml had a primary sequence highly divergent from the sequences of those that had an MIC <4µg/ml is similar to what has been found in *Streptococcus pneumoniae*, except that the *pbp1, pbp2b,* and *pbp2x* sequence in resistant clinical isolates revealed mosaic genes, with blocks of sequences that differed up to 23% (Zapun et al., 2008). To determine if the 5% variation and division of isolates into two groups (Figure 5) was restricted to *pbp5*, or if this was a difference that could be found on a larger genomic scale, we looked at three other genes, including *pbp2*, encoding proteins thought to be on the cell surface and, thus, possibly more likely to change compared to more conserved housekeeping genes. The sequences of *gls20*, *wlcA*, and *pbp2* from the 11 sequenced USA isolates available on the Broad Institute and NCBI were analyzed (Figure 6). The isolates also fell into two distinct lineages based on each of the other three genes, with up to a 3% nucleotide difference for *gls20*, up to a 7% difference for *pbp2*, and up to a 10% difference for *wlcA* between the two groups. These two lineages generally corresponded with the *pbp5*-S and *pbp5*-R groups, which also corresponded to a CC17 genogroup (ST17, plus single locus variants, and/or double locus variants of it) and a non-CC17 genogroup. While most of the time the isolates in the *pbp5*-S and *pbp5*-R groups also grouped together using the other three genes, this was not always the case (Figure 6). For example, Efm 1231501 groups with the *pbp5*-R group using *pbp2* and *wlcA* genes, despite being a non-CC17 isolate, and with the *pbp5*-S group for *pbp5* and *gls20*. Efm 1231408 grouped with the *pbp5*-S group using *wlcA*, even though it is a CC17 genogroup isolate, and with the *pbp5*-R for *pbp5, pbp2* and *gls20*. (Galloway-Pena et al., 2011)

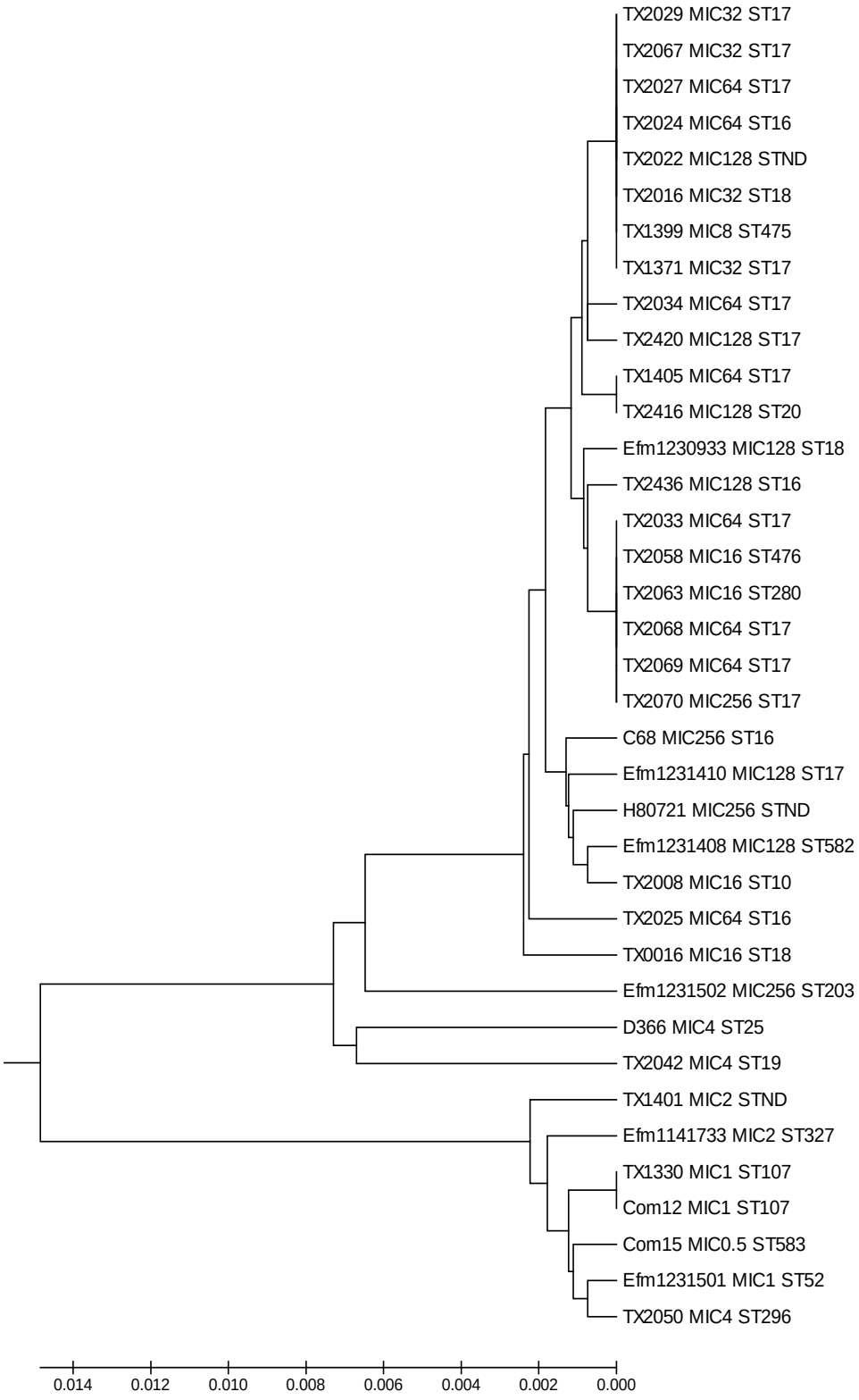


FIGURE 5. PBP5 phylogeny coincides with MIC. A UMPGA tree of PBP5 was constructed using the amino acid sequence from all strains analyzed in this study. PBP5 splits isolates into

two groups using the amino acid sequence and these two groups coincide with the MIC. Isolates with an ampicillin MIC < 4µg/ml all group together in the lower branch of the tree (*pbp5*-S group) (except for TX2050, which has an MIC = 4µg/ml), while all isolates with an MIC >4µg/ml group together in the top branch of the tree (*pbp5*-R group). The minimal inhibitory concentration in micrograms per milliliter and multi-locus sequence type is listed next to the strain name. STND designates an ST that was not done. Figure from (Galloway-Pena et al., 2011).

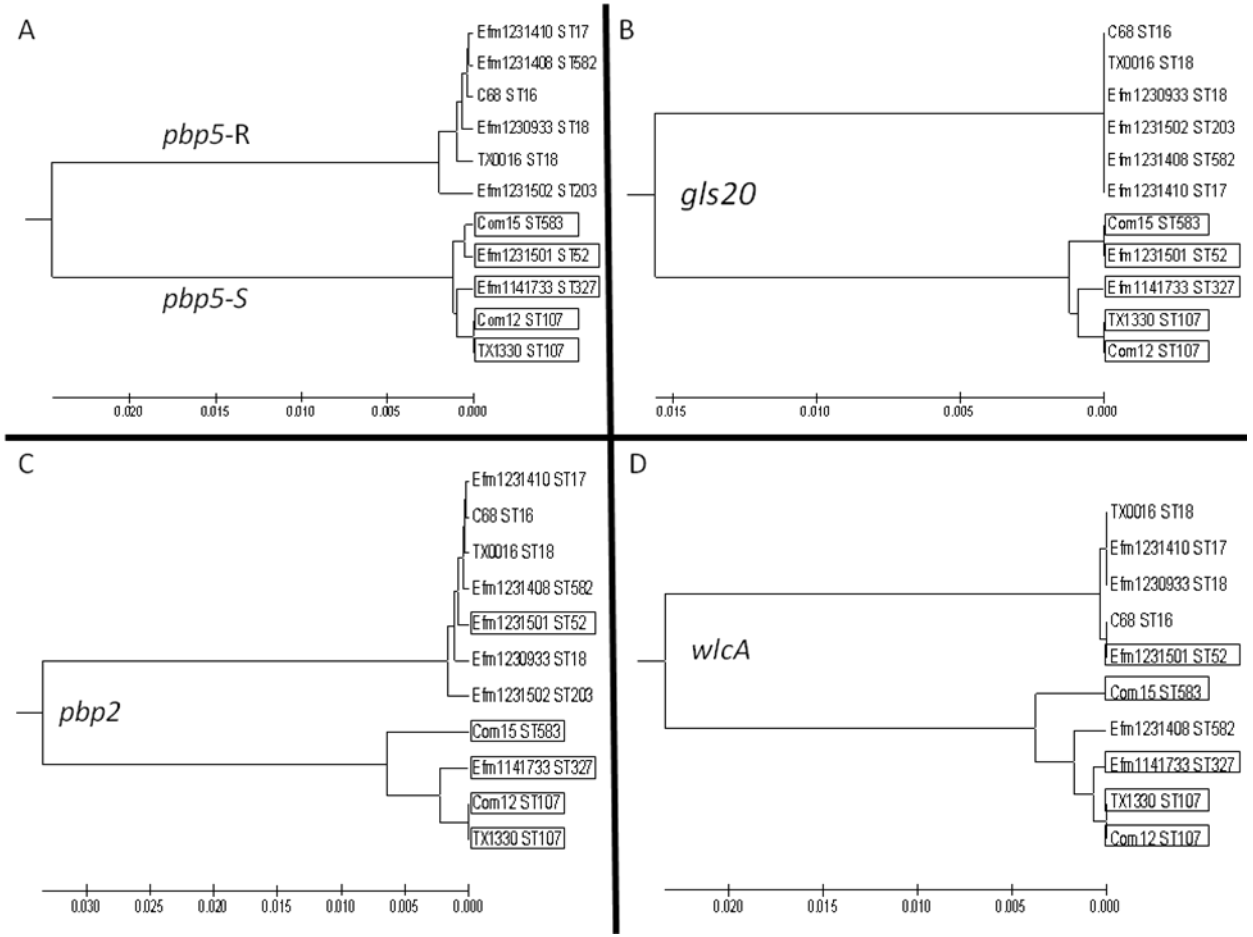


FIGURE 6. UPGMA trees of four *E. faecium* genes shows isolates generally fall into two distinct lineages.

UPGMA trees of *pbp5* (2A), *gls20* (2B), *pbp2* (2C), and *wlcA* (2D) are depicted using the sequenced USA isolates and available on the Broad Institute or NCBI. In order to determine if other genes showed the same two groupings as *pbp5* (designated *pbp5*-S and

pbp5-R) (2A), UPGMA trees of *gls20* (2B), *pbp2* (2C), and *wlcA* (2C) were made. The boxes indicate isolates that share the ampicillin *pbp5*-S (MIC < 4µg/ml) consensus. The sequence for *wlcA* of Efm1231502 was unavailable. Figure from (Galloway-Pena et al., 2011).

[2.4] CHAPTER 1 DISCUSSION

ST17 isolates were clearly present in United States hospitals by 1982 (Table 4 and Figure 1), predating the earliest published CC17 genogroup isolate (recovered in Great Britain) by seven years and predating the previously earliest known United States CC17 related isolate (DO/ TX16) by nine years (http://efaecium. mlst.net). The key characteristics of these early United States CC17 genogroup isolates were 1) ampicillin resistance, 2) high-level resistance to streptomycin, and 3) the presence of the *fms* genes (Table 3, Table 4, and Figure 1).

In addition to the earliest CC17 outbreak isolates, all the United States isolates within the CC17 genogroup analyzed were found to contain similar *fms* gene profiles with all isolates harboring at least 13/14 *fms* genes (either *fms20* or *fms18* missing). This result was clearly different from non-CC17 United States isolates, which were found to contain only five to nine *fms* genes (*P*-values ranging from 0.001 to 0.029 except for *fms20*) (Table 4 and Figure 3).

Although esp_{Efm} was present in one of the earliest ST17 isolates (TX2038 isolated in 1982), it was not found in most of the early outbreak isolates analyzed. However, its presence in CC17 isolates versus non-CC17 isolates was statistically significant ($P = 0.0002$) (Figure 2). This finding suggests that esp_{Efm}, was likely acquired at different times during *E. faecium's* hospital-adaptation (most likely through horizontal transfer). The fact that *fms* genes were present earlier raises the possibility that these pili and adhesins played more of a role in *E. faecium's* early adaptation to becoming a more prominent opportunistic pathogen then esp_{Efm}, at least in the United States.

Additional testing for the hyl_{Efm} gene indicates that its presence ($P = 0.001$ for CC17 versus non-CC17) increased around the late 1980's (Table 4, Figure 1, and Figure 2). In addition, vancomycin resistance ($P = 0.044$ for CC17 versus non-CC17) first appeared in 1990

and was due to the acquisition of the *vanA* gene cluster among the CC17 genogroup United States isolates (Table 3, Figure 1, and Figure 2), supporting the emergence of vancomycin resistance occurring after ampicillin resistance. Therefore, one might hypothesize that early United States ampicillin resistant hospital-associated isolates carrying *fms* genes, subsequently acquired esp_{Efm} and/or the hyl_{Efm}-containing plasmid, which can also contain vancomycin and gentamicin resistance determinants. The sum of these attributes likely contributed to the recent success of *E. faecium* as a nosocomial pathogen.

Comparable findings were found when a similar analysis was performed on South American VREfm isolates. Although VREfm prevalence in South America is still low compared to the United States, the VREfm isolates have genotypic and phenotypic characteristics similar to the United States isolates studied: (i) South American VREfm are mostly CC17 related isolates (mostly belonging to ST412 and ST18) and (ii) a high percentage of the South American CC17 related VREfm isolates contain MSCRAMM genes (*fms* genes) and hyl_{Efm} (Table 5 and Figure 4).

One observation of my collaborators was that in some of the South American isolates, the *fms20-fms21* cluster existed on the same transferrable plasmid as the hyl_{Efm} gene, which often contains *vanA*, conferring vancomycin resistance, and *aac(6')-aph(2")*, conferring gentamicin resistance (Kim et al., 2010; Panesso et al., 2010) Interestingly, almost all of the VREfm United States isolate studies also exhibited high-level resistance to gentamicin and harbored the hyl_{Efm} gene (Table 3 and 4). These similarities in results suggest that the recent population genetics of VREfm in northern South America reflects the early isolates responsible for hospital outbreaks in the United States. Due to this finding, one might predict an increase in the incidence of infections caused by CC17-related *E. faecium* in South America over the

next decade as this clone continues to expand and evolve by the horizontal transfer of plasmid associated virulence factors and antibiotic resistance determinants.

Although ampicillin resistance was present significantly more often in United States CC17 isolates versus non-CC17 isolates ($P = 0.0007$), 8 out of the 41 United States ampicillin resistant isolates were not STs related to CC17, showing that not all ampicillin resistant isolates are closely related (Table 3, Figure 1, and Figure 2). *E. faecium* isolates intrinsically resistant to β-lactam antibiotics (Arias and Murray, 2008) were sporadically reported as early as 1965 (Willems et al., 2005). However, reports of highly ampicillin resistant *E. faecium* isolates causing infections did not occur until the 1980's (Boyce et al., 1992; Bush et al., 1989; Coudron et al., 1984; Eliopoulos et al., 1982; Sapico et al., 1989). The hypothesis that the presence of ampicillin resistance was one of the first steps of hospital-adaptation by *E. faecium* is supported by the finding that ampicillin resistance existed in the earliest CC17 outbreak isolates (Leavis et al., 2006a). Only two of the United States CC17 related isolates analyzed in this study were found to be ampicillin susceptible according to CLSI guidelines (MIC < 16μg/ml).

Furthermore, it was found that ampicillin MICs incrementally increased over time within the United States CC17 related isolates (Table 3, Figure 1, and Table 8). This finding suggested further mutations or overexpression of the ancestral *pbp5* gene might have evolved in the hospital-adapted isolates, decreasing the susceptibility to ampicillin. Previous publications suggested that particular substitutions present in combination could amplify resistance levels (Poeta et al., 2007; Rice et al., 2004) as well as the ability of clinical isolates to transfer a low-affinity PBP5 (Rice et al., 2005). Therefore, we set out to determine if either of these scenarios was true.

Although there is a PBP5 consensus sequence among isolates with higher ampicillin MICs (Table 8), the increases in MIC seen among strains with the *pbp5*-R sequences could not be explained using sequence correlation. Thus, other factors such as regulation of expression or translational modifications are likely to be involved. But, overall the results suggest that the presence of a basic consensus sequence of a more resistant PBP5 that is different from the ancestral PBP5 is needed to obtain high MICs of ampicillin in hospital-associated *E. faecium* (Table 7, and Figure 5).

We also observed the same genomic variability seen in the *pbp5* genes in three other genes studied, suggesting that strain background might be a factor in the ability to obtain higher MICs or cause infections. The observation that *pbp5*, *pbp2*, *gls20*, and *wlcA* have sequences that phylogenetically branch into two similar groups, suggested that there are two distinct evolutionary lineages of *E. faecium* (Figure 5 and 6). Moreover, the data showing that switching occurred between groups depending on the gene (i.e., that the other genes were not always associated with the same group as seen with the *pbp5* gene) (Figure 5 and 6), suggests a horizontal exchange of genes between individual isolates of the two lineages. The hypothesis of two distinct *E. faecium* clades, one community-associated and another hospital-associated, will be further discussed in Chapter 2.

[3] CHAPTER 2

Genomic Analyses Demonstrate a Distant Separation of the Hospital and Community-Associated Clades of *Enterococcus faecium*

[3.1] CHAPTER 2 INTRODUCTION

Early population biology studies suggested that clinical *E. faecium* strains belonged to a genetic lineage named CC17 (Leavis et al., 2006a; Top et al., 2008; Willems et al., 2011) that is distinct from non-clinically associated clonal groups. However, studies using Bayesian modeling and other methods subsequently reported that eBURST-based clustering is inaccurate for determining evolutionary decent for species with high levels of recombination (like *E. faecium*) (Didelot and Falush, 2007; Tang et al., 2009; Turner et al., 2007; Willems and van Schaik, 2009). Even so, other studies using mixed whole-genome array, AFLP, and pyrosequencing still indicated the existence of two different host-specific subpopulations, (distinguished by antibiotic resistance, distinct STs, and the accessory genome) in which one subgroup consists primarily of hospital-associated strains and the other community-associated strains (Leavis et al., 2007; van Schaik et al., 2010; van Schaik and Willems, 2010).

Many publications argue that the accessory genome of hospital-associated strains is the driving force behind the evolution of this organism, rather than evolutionary descent (Leavis et al., 2007; van Schaik et al., 2010). However, underlying differences in the core genome were indicated by pyrosequencing (van Schaik et al., 2010) and by the results discussed in the last chapter where strains separated into two divergent groups based on the genes *gls20*, *pbp5*, *pbp2*, and *wlcA* (Galloway-Pena et al., 2011). So, even though mobile genetic elements, antibiotic resistance, and virulence determinants may well contribute to persistence in the

hospital environment, the data also suggest there might be fundamental differences between the two subpopulations at the level of the core genome that could be the foundation for *E. faecium's* success as a nosocomial pathogen.

This chapter discusses data from two published studies. The first study (Galloway-Pena et al., 2012) reflects on the recent studies discussed above which pointed to the existence of two subpopulations of *E. faecium*, one containing primarily commensal/community-associated (CA) strains and one containing mostly clinical or hospital-associated (HA) strains. As discussed in the previous chapter, we previously found a 3-10% difference between four genes from HA-clade strains vs. CA-clade strains, including 5% difference between *pbp5-R* allele of HA strains and *pbp5-S* allele of CA strains, further supporting the two clade concept. In order to further investigate the differences in the core genome of these subpopulations, we analyzed 100 genes from 21 *E. faecium* genome sequences. Concatenated sequences, SNPs, and individual gene analysis of the 100 genes all identified two distinct groups. Concatenated sequences differed 3.5-4.2% between the two clades. Molecular clock calculations using sSNP analysis suggest that the two clades diverged ~300,000 to $\geq$ 1 million years ago, depending on the mutation rate used. These data confirm the existence of two clades of *E. faecium* at the core genomic level and show that their divergence was a distant evolutionary event.

The second study (Qin et al., 2012) includes selected data from our report of the complete genome sequence of the *E. faecium* strain TX16. Whole genome comparison, phylogenetic, and gene content similarity analyses of TX16 with 21 *E. faecium* draft genomes confirmed the CA and HA clades found through the 100 core gene analysis. This study also revealed many mobile genomic loci unique to the HA clade, such as IS elements and transposons. It was also found that antibiotic resistance genes are enriched in HA-clade strains

whereas CRISPRs (Clustered Regularly Interspaced Short Palindromic Repeats) are enriched in the CA-clade.

[3.2] CHAPTER 2 MATERIALS AND METHODS

[3.2.1] Selection of 100 orthologs.

Genome sequences of 21 *E. faecium* available from NCBI were studied (Table 9). To investigate whether there is a clear separation at the genome level into distinct groups, we selected 100 orthologs (Table 10). These 100 orthologs were selected based on position (spread over different regions of the chromosome) and their presence in all strains as housekeeping genes or putative non-antigenic genes, including ribosomal proteins (Figure 7). Ortholog groups of *E. faecium* genomes were identified using OrthoMCL program using BLASTP E value of 1e-5 and default MCL inflation parameter of 1.5 with 80% sequence identity and 60% match length cutoffs. However, only those genes with the same size in base pairs were chosen. (Galloway-Pena et al., 2012)

[3.2.2] Comparative DNA sequence analysis.

The 100 chosen orthologs nucleotide sequences were concatenated into one continuous sequence for each of the 21 *E. faecium* strains and a pair wise analysis using the Poisson correction method (Zuckerkandl, 1965) on MEGA 4.0.2 software was performed (Kumar et al., 2008; Tamura et al., 2007). UPGMA phylogenetic trees (Sneath and Sokal, 1962) were constructed using the ClustalW alignment of the concatenated sequence (Chenna et al., 2003; Larkin et al., 2007; Thompson et al., 1994). The divergence score was calculated by taking the distance of the branch lengths between two strains, divided by the total distance (or sum of all the branch lengths) and multiplying by 100 (Kumar et al., 2008). In addition, each individual gene for the 21 genomes was also analyzed separately using ClustalW and UPGMA trees were generated using MEGA 4.0.2. software and analyzed to see which branch it segregated to (community or hospital clade) (Chenna et al., 2003; Kumar et al., 2008; Tamura et al., 2007; Thompson et al., 1994). If there were not two distinct branches, the gene was excluded from this analysis (8 of the 100 were excluded, leaving 92 genes). (Galloway-Pena et al., 2012)

TABLE 9. The sequence type, country, date, and source of isolation for the 21 sequenced *E. faecium* genomes used in this study.

Strain	ST[a]	Country[b]	year	Source	Accession #
C68	16	USA (OH)	1998	Endocarditis patient (blood)	ACJQ00000000
Com12	107	USA	2006	Healthy volunteer fecal sample	ACBC00000000
Com15	583	USA	2006	Healthy volunteer fecal sample	ACBD00000000
D344SRF	21	France	1985	Clinical isolate	ACZZ00000000
E980	94	Netherlands	1998	Healthy volunteer fecal sample	ABQA01000001
E1039	42	Netherlands	1998	Healthy volunteer fecal sample	ACOS00000000
E1071	32	Netherlands	2000	Hospitalized patient fecal sample	ABQI01000001
E1162	17	France	1997	Blood Culture of Hospitalized Patient	ABQJ00000000
E1636	106	Netherlands	1961	Blood Culture of Hospitalized Patient	ABRY01000001
E1679	114	Brazil	1998	Swab of Vascular Catheter Tip	ABSC01000001
TX16	18	USA (TX)	1992	Endocarditis patient (blood)	ACIY00000000
TX82	17	USA (TX)	1999	Endocarditis patient (blood)	AEBU00000000
TX0133	17	USA (TX)	2006	Endocarditis patient (blood)	AECH00000000
TX1330	107	USA	1994	Healthy volunteer fecal sample	ACHL00000000
U0317	78	Netherlands	2005	UTI of Hospitalized Patient (Urine)	ABSW01000001
1141733	327	unk.	unk.	Blood Culture of Hospitalized Patient	ACAZ00000000
1230933	18	unk.	unk.	Wound Swab of Hospitalized Patient	ACAS00000000
1231408	582	unk.	unk.	Blood Culture of Hospitalized Patient	ACBB00000000
1231410	17	unk.	unk.	Skin and Soft Tissue Infection	ACBA00000000
1231501	52	unk.	unk.	Blood Culture of Hospitalized Patient	ACAY00000000
1231502	203	unk.	unk.	Blood Culture of Hospitalized Patient	ACAX00000000

[a] ST is the sequence type by multilocus sequence typing

[b] "unk." means the information for this isolate is unknown

Table from (Galloway-Pena et al., 2012)

TABLE 10. 100 core genes chosen for analysis and concatenation in this study.

ORF	Start[d]	End	Strand	Size (bp)	Name	Description
HMPREF0351_10005[c]	4312	6258	+	1947	*gyrB*	DNA topoisomerase subunit B
HMPREF0351_10009[b]	9872	10108	+	237	*rpsR*	ribosomal protein S18
HMPREF0351_10017	20752	22482	-	1731	*poxB*	pyruvate oxidase
HMPREF0351_10054	58163	60250	+	2088	*fusA*	protein-synthesizing GTPase
HMPREF0351_10068[a,b]	67123	67491	+	369	*rplN*	ribosomal protein L14
HMPREF0351_10078[b]	71325	72620	+	1296	*secY*	preprotein translocase subunit SecY
HMPREF0351_10083[b]	74767	75705	+	939	*rpoA*	RNA polymerase subunit alpha
HMPREF0351_10125[b]	123688	123909	+	222	*metG*	methionine--tRNA ligase
HMPREF0351_10135	131070	133715	+	2646	*mutS*	DNA mismatch repair protein MutS
HMPREF0351_10137	135863	136423	+	561	*maf*	septum formation protein Maf
HMPREF0351_10164	162881	164251	-	1371	*dinF*	MATE family protein
HMPREF0351_10231[b]	229498	229698	+	201	*rpmI*	ribosomal protein L35
HMPREF0351_10244[a,c]	240647	241726	+	1080	*pepA*	M42 family glutamyl aminopeptidase
HMPREF0351_10375	354375	355688	+	1314	*yieG*	NCS2 family purine:cation symporter
HMPREF0351_10389[c]	374025	374741	+	717	*glpF2*	MIP family channel protein
HMPREF0351_10406	389976	390635	+	660	*rcfA*	CRP/FNR transcriptional regulator
HMPREF0351_10439	429479	430207	+	729		integral membrane protein
HMPREF0351_10460	453947	454675	-	729	*agrA*	accessory gene regulator protein A
HMPREF0351_10474[b]	471998	472561	+	564	*efp*	elongation factor P
HMPREF0351_10476	473048	474307	+	1260	*lysA*	diaminopimelate decarboxylase
HMPREF0351_10493	491476	492381	+	906		AraC family transcriptional regulator
HMPREF0351_10496	493430	494326	+	897		LysM family surface protein
HMPREF0351_10540	536179	537375	-	1197	*telA*	tellurite resistance protein
HMPREF0351_10568	567155	568342	+	1188	*ftsW*	cell division membrane protein
HMPREF0351_10578[a]	575515	575985	+	471	*iscU*	Fe-S cluster formation protein
HMPREF0351_10611	612012	612923	+	912	*ppiA*	peptidylprolyl isomerase
HMPREF0351_10633	631143	631583	+	441	*ntd*	nucleoside deoxyribosyltransferase
HMPREF0351_10659	664118	664564	-	447	*nrdI*	ribonucleotide reductase
HMPREF0351_10683	684086	685021	+	936	*murB*	N-acetylmuramate dehydrogenase
HMPREF0351_10719	723392	725002	+	1611	*pyrG*	CTP synthase
HMPREF0351_10743[b]	745909	746256	+	348	*yabA*	regulator of replication initiation
HMPREF0351_10765[b]	770338	770646	+	309	*rplU*	ribosomal protein L21
HMPREF0351_10791	791519	793711	+	2193	*pbpA*	penicillin-binding protein 1
HMPREF0351_10798	801185	801862	+	678	*alr2*	alanine racemase

HMPREF0351_10801	803629	804336	+	708	*divIVA*	cell division initiation protein DivIVA
HMPREF0351_10810	813420	816122	+	2703	*sigL*	RNA polymerase sigma subunit
HMPREF0351_10903[a,c]	877677	878897	+	1221	*tagH*	teichoic acid ABC superfamily
HMPREF0351_10938	914820	915452	+	633		LuxR family response regulator
HMPREF0351_10945	920237	920845	+	609	*sodA*	superoxide dismutase
HMPREF0351_10961[b]	937960	938226	+	267	*ptsH*	PTS family porter component HPr
HMPREF0351_10976	950670	952049	+	1380	*accC*	biotin carboxylase
HMPREF0351_11002	982334	984118	+	1785	*pyk*	pyruvate kinase
HMPREF0351_11005[b]	985868	986047	+	180	*rpmF*	ribosomal protein L32
HMPREF0351_11081[c]	1066164	1067351	+	1188	*tufA2*	elongation factor EF1A
HMPREF0351_11103	1091135	1091356	-	222	*fer*	ferredoxin
HMPREF0351_11142	1127033	1127824	+	792	*codY*	CodY family transcriptional regulator
HMPREF0351_11157	1143520	1144467	+	948		LacI family transcriptional regulator
HMPREF0351_11204	1187269	1188564	+	1296	*purB*	adenylosuccinate lyase
HMPREF0351_11221[a]	1205721	1207247	+	1527	*araG*	L-arabinose ABC superfamily
HMPREF0351_11245	1232977	1233516	+	540	*comEB*	competence protein ComEB
HMPREF0351_11282	1267078	1268121	+	1044	*pheS2*	phenylalanine--tRNA ligase
HMPREF0351_11311	1296129	1298126	+	1998	*tktA*	transketolase
HMPREF0351_11315[c]	1300562	1301008	+	447	*fur2*	Fur family transcriptional regulator
HMPREF0351_11316	1301281	1302633	+	1353	*nox*	NADH dehydrogenase
HMPREF0351_11335	1321937	1323286	+	1350	*rhlB*	ATP-dependent RNA helicase
HMPREF0351_11369[a]	1356073	1356927	+	855	*aroE*	shikimate dehydrogenase
HMPREF0351_11548	1517751	1520360	-	2610	*clpB*	S14 family endopeptidase Clp
HMPREF0351_11556	1526794	1527720	-	927	*dnaI*	primosomal protein DnaI
HMPREF0351_11568	1538758	1540938	-	2181	*tex*	S1 domain RNA-binding protein
HMPREF0351_11610	1583663	1585222	+	1560	*eriC*	CPA2 family cation:proton antiporter
HMPREF0351_11619[a,c]	1593185	1594474	-	1290	*tig*	cell division trigger factor
HMPREF0351_11642	1618069	1618800	-	732	*racX*	aspartate racemase
HMPREF0351_11679	1657393	1659102	-	1710	*proS*	proline--tRNA ligase
HMPREF0351_11694	1673135	1674364	-	1230	*arcA*	arginine deiminase
HMPREF0351_11718[b]	1693974	1695869	-	1896	*asnB*	asparagine synthase
HMPREF0351_11738	1712076	1713422	-	1347	*pgi*	glucose-6-phosphate isomerase
HMPREF0351_11759	1733232	1733942	-	711	*alsD*	acetolactate decarboxylase
HMPREF0351_11836[b]	1805300	1805575	-	276	*rpsP*	ribosomal protein S16
HMPREF0351_11857[c]	1824675	1825649	-	975	*cbh3*	choloylglycine hydrolase
HMPREF0351_11894	1855186	1856484	-	1299	*asnS*	asparagine--tRNA ligase
HMPREF0351_11904[a]	1864309	1865616	+	1308	*dacA*	S11 D-Ala-D-Ala carboxypeptidase
HMPREF0351_11979[c]	1928482	1929453	-	972	*phoH*	phosphate starvation-inducible

HMPREF0351_12018[c]	1972203	1972880	-	678	*phoU*	phosphate regulatory protein
HMPREF0351_12030	1984852	1985403	-	552	*yfiA*	ribosome-associated inhibitor
HMPREF0351_12057	2005959	2006195	-	237	*secG*	Sec family Type I secretory protein
HMPREF0351_12066	2013386	2014576	-	1191	*pgk*	phosphoglycerate kinase
HMPREF0351_12075[c]	2021207	2021827	-	621	*rpoE2*	RNA polymerase subunit delta
HMPREF0351_12120	2067231	2068871	+	1641	*sfcA*	malate dehydrogenase
HMPREF0351_12172	2114033	2114770	-	738	*glpQ2*	glycerophosphodiesterase
HMPREF0351_12186	2126586	2129231	-	2646	*valS*	valine--tRNA ligase
HMPREF0351_12209	2155856	2156365	+	510	*pgpA*	phosphatidylglycerophosphatase
HMPREF0351_12250	2197932	2199041	+	1110	*msmK*	maltose ABC superfamily
HMPREF0351_12258[c]	2205177	2207186	-	2010	*metG2*	methionine--tRNA ligase
HMPREF0351_12275	2222997	2224433	-	1437	*bglB2*	beta-glucosidase
HMPREF0351_12305	2246582	2247934	+	1353	*murE*	UDP- ligase
HMPREF0351_12332	2275213	2276637	-	1425	*araA*	L-arabinose isomerase
HMPREF0351_12354[c]	2300216	2300500	-	285	*groES*	chaperone GroES
HMPREF0351_12371	2316689	2317333	-	645		GntR family transcriptional regulator
HMPREF0351_12412	2352668	2353534	-	867	*rrmA*	23S rRNA methyltransferase A
HMPREF0351_12423	2364403	2365056	-	654	*mecA*	competence negative regulator MecA
HMPREF0351_12437	2384535	2385794	-	1260	*cypX*	cytochrome P450
HMPREF0351_12458	2401433	2402482	-	1050	*comGB*	competence protein ComGB
HMPREF0351_12497	2452371	2453228	-	858	*cscK3*	fructokinase
HMPREF0351_12540	2508096	2509289	-	1194	*metK*	methionine adenosyltransferase
HMPREF0351_12545	2520662	2522773	-	2112	*ftsH*	M41 family endopeptidase FtsH
HMPREF0351_12572	2552543	2553157	-	615	*gmk*	guanylate kinase
HMPREF0351_12605[c]	2587423	2588580	-	1158		PilT domain protein
HMPREF0351_12650	2625999	2626466	-	468	*dps*	DNA-binding protein Dps
HMPREF0351_12682	2667041	2668387	-	1347	*gor2*	glutathione-disulfide reductase
HMPREF0351_12694	2688720	2689610	-	891	*spo0J*	stage 0 DNA-binding protein spoOJ

[a] Genes left out of the individual gene analysis (i.e. genes that did not show clade distinctions)

[b] Genes that do not have non-synonymous changes in their encoded protein

[c] Genes that have non-synonymous changes in their encoded protein, but are not clade specific

[d] Refers to the nucleotide start and end sites on the DO (TX16) chromosome

[e] An empty cell indicates a KEGG number was not identified

Table from (Galloway-Pena et al., 2012)

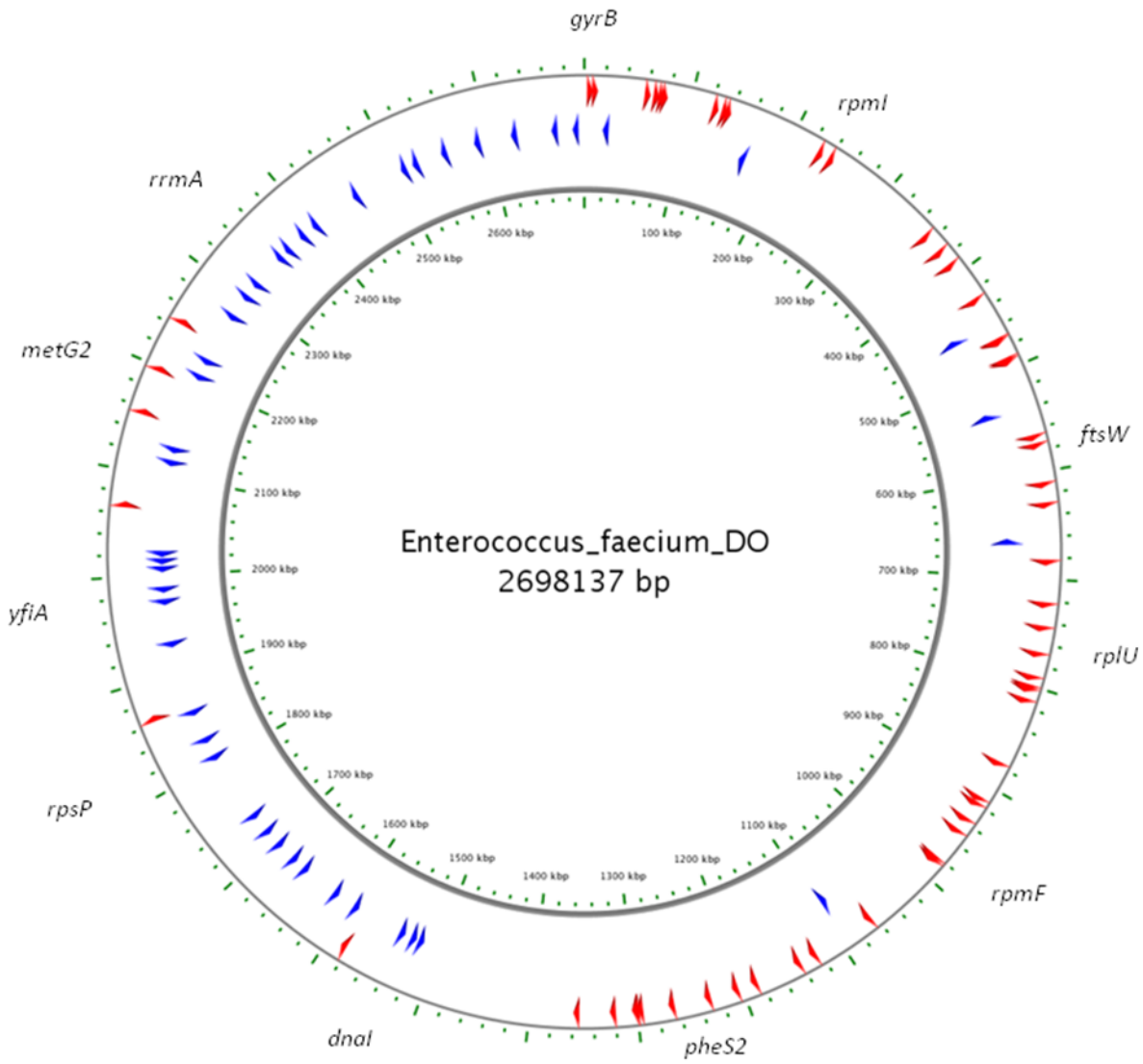


FIGURE 7. Position of the 100 genes analyzed on the TX16 (DO) chromosome. The chromosomal map of the *Enterococcus faecium* genome of the endocarditis isolate TX16, commonly referred to as DO. The arrows represent the 100 genes chosen for clade and SNP analysis and their position on the chromosome. The red arrows depict genes transcribed in the forward orientations, and blue arrows depict genes transcribed in the reverse orientation. The general locations of a subset of the 100 genes are labeled. Figure from (Galloway-Pena et al., 2012)

[3.2.3] SNP analysis.

To further investigate the differences among *E. faecium* strains, all SNP differences were extracted from the aligned 21 concatenated sequences and were concatenated into one continuous DNA sequence for each strain, compared to each other for nucleotide identity and divergence, and an UPGMA tree was constructed using the same methodology stated previously. We also calculated the number of SNPs that were strain specific (defined as found in only one strain) versus those found in two or more strains. The number of non-synonymous changes was analyzed and it was determined in which genes they existed. Strain specific amino acid changes were excluded. We then identified clade-specific non-synonymous changes, defined as those changes that were present in at least two community-associated *E. faecium* strains but not in *E. faecium* strains within the HA clade, or present in 14 of the HA strains but not in the community strains. (Galloway-Pena et al., 2012)

[3.2.4] Molecular clock estimation.

The molecular clock estimation requires four components, the number of synonymous SNPs, the number of potential SNP sites, the mutation rate, and the number of generations per year (Foster et al., 2009; Lenski et al., 2003; Van Ert et al., 2007). In order to estimate the rate of evolution of the two subpopulations, the number of synonymous SNPs (sSNPs) from the 100 gene analysis determined above was used. The potential sSNP sites were calculated by finding all three-base codons that could be used within the 100 genes and adding together all the sSNP sites from each codon. The median number of sSNPs between the strains within a clade was used to calculate the molecular clock for strains within a clade. Similarly, the median number of sSNPs for each clade was used to compare between the two clades. Since the synonymous mutation rate for *E. faecium* is unknown, a synonymous mutation rate of 1.4×10^{-10} mutations per base pair per generation based on mutation rates from *Escherichia coli* was used (Foster et al., 2009; Lenski et al., 2003). The number of generations per year of *Enterococcus* species in the host or in the environment is also unknown, so a range of possible generation times 100, 200, and 300 was used, also based on *E. coli* (Foster et al., 2009; Lenski et al., 2003). Estimates for a closer relative *Bacillus anthracis* are approximately 43 generations per year and its mutation rate is 5.2×10^{-10}, and this higher mutation rate was also used (Foster et al., 2009; Lenski et al., 2003; Van Ert et al., 2007). The following equation was used to determine the time of divergence for each comparison: the number of sSNP/ (the number of possible sSNP sites x mutation rate x the number of generations per year x 2). The "2" in the denominator of the equation is used to account for the time of divergence of the two genomes, or the two groups compared (Achtman et al., 2004; Foster et al., 2009; Van Ert et al., 2007). (Galloway-Pena et al., 2012)

[3.2.5] Analysis of the 16S rRNA and estimation of the time of divergence.

The 16S rRNA gene was sequenced using the universal 16S rRNA primers B27F, 785F, 805R, and primers designed for outside the 16S rRNA gene for *E. faecium*, 16SEfmOS F1 5'-ATCGCAAGATTGTTCGAAC -3', and 16S EfmOSR2 5'-CTTAGAAAGGAGGTGATCCAG -3'. The entire 16S rRNA of TX16 (DO) and TX1330, as representatives of the two clades, were resequenced. Sequences were

extracted from all other strains using the NCBI sequence. Strains that had an incomplete 16S rRNA sequence (E1636, 1231502, Com12, Com15, 1230933, TX0133A, E1039, and TX82) were also resequenced. The sequences were aligned and compared using MEGA 4.0.2 software. The SNPs and percent nucleotide difference between the strains were determined, and the determination of time for divergence was based on a 1% change per 25 million years ago for 16S rRNA (Ochman and Wilson, 1987). (Galloway-Pena et al., 2012)

[3.2.6] Genome characterization.

BLASTN and BLASTX as well as ISfinder server (Varani et al., 2011) were used to identify IS sequences and transposons in the TX16 chromosome and plasmids. Genomic regions with homology to IS and transposon sequences from both BLAST analyses were verified with the gene annotation of TX16. Both BLAST searches identified many small regions as a part of IS elements and transposons. Regions with shorter than 60% match length to reference sequences were excluded from further analysis. Identified genes/regions by analyses above were also used to perform the BLAST search against the other 21 *E. faecium* genomes to investigate whether there are clade specific presences or absences. (Qin et al., 2012)

Six genomes that had yet to be studied for CRISPR-loci were analyzed for CRISPR loci (TX1330, TX16, TX82, TX0133A, D344SRF, and C68). We searched for CRISPR loci in the six genomes by performing BLAST using the sequences from the ORFs previously described for CRISPR-loci in *E. faecium* EFVG_01551 to EFVG_01555 (Palmer and Gilmore, 2010), as well as using CRISPRfinder (http://crispr.u-psud.fr/Server/CRISPRfinder.php) and the CRT program (Bland et al., 2007) to detect prophage CRISPR palindromic repeats in TX16. (Qin et al., 2012)

[3.2.7] Ortholog, phylogenetic and multi-locus sequence typing (MLST) analysis.

Protein ortholog groups of *E. faecium* genomes were identified using OrthoMCL program (Li et al., 2003) using BLASTP E value of 1e-5 and default MCL inflation parameter of 1.5 with 80% sequence identity and 60% match length cutoffs. The match length percentage was set relatively low because all the genomes except TX16 are draft sequences. The dissimilarity in gene content among the *E. faecium* genomes was calculated using Jaccard distance (1- Jaccard coefficient) as described previously (Suzuki et al., 2011), and the Jaccard distance matrix was used for hierarchical clustering using the unweighted pair group method with arithmetic mean (UPGMA). Single-copy orthologs with the same length in all strains were chosen for phylogenetic analysis after removing genes that may have undergone recombination detected by PHI program (Bruen et al., 2006). Multiple sequence alignments were performed by MAFFT program (Katoh et al., 2002) and the topology of the phylogenetic tree was inferred by maximum-likelihood algorithm using PhyML (Guindon et al., 2010) with bootstrap value of 100. 16S rRNA phylogenetic analysis was performed in another manuscript (Galloway-Pena et al., 2012). iTOL program (Letunic and Bork, 2007) was used for phylogenetic tree visualization. (Qin et al., 2012)

The *in silico* multi-locus sequence types were determined either by extracting the allele types of *adk*, *atpA*, *ddl*, *gdh*, *gyd*, *pstS*, and *purK* from the genomic sequence, or using the allele numbers previously obtained through experimentation (Galloway-Pena et al., 2011). The allele numbers and sequence types were used to construct an UPGMA dendogram using S.T.A.R.T.2 software (http://pubmlst.org/). (Qin et al., 2012)

[3.2.8] Identification of putative virulence-associated genes and antibiotic resistance determinants.

Putative virulence genes were identified by BLASTP of *E. faecium* ORF protein sequences to the enterococcal virulence factors in the Virulence Factors Database (VFDB) (Chen et al., 2005), and hits were manually inspected. (Qin et al., 2012)
To identify antibiotic resistance genes, BLASTN was performed using the nucleotide sequences of 13 antibiotic resistance genes including *cat* (chloramphenicol O-acetyltransferase) using the EfmE1071_2206 sequence which is an ortholog to the *cat* gene found on the *E. faecium* plasmid pRUM (Grady and Hayes, 2003) *ermA* (rRNA adenine N-6-methyltransferase) using the EfmE1679_0214 sequence and located on Tn554 (Murphy et al., 1985); *ermB* (rRNA adenine N-6-methyltransferase) using the EfmE1071_2296 sequence, an ortholog to the *ermB* gene found on the *E. faecalis* plasmids pRE25 and pSL1(Schwarz et al., 2001); *aad6* (aminoglycoside 6-adenylyltransferase) using the EfmE1071_1021 sequence an ortholog to the genes found on the *E. faecalis* plasmid pEF418 (Genbank:AF408195); *aad9* (streptomycin 3''-adenylyltransferase) using EfmE1679_0213 sequence and located on Tn*554* (Murphy et al., 1985); *aadE* (aminoglycoside 6-adenylyltransferase) using EfmU0317_2169 sequence an ortholog to the gene found on the *E. faecalis* plasmid pRE25 (Schwarz et al., 2001); *aacA-aphD* (bifunctional aminoglycoside modifying enzyme) using the EfmU0317_2161 sequence; *tetL* using the EfmE1071_1017 sequence (Burdett et al., 1982); *tetM* using the EfmE1162_0404 sequence (Burdett et al., 1982); *vanA* using the EfmE1071_0104 to EfmE1071_0110 sequence which is identical to the *vanA* gene cluster found on Tn*1546* (Arthur et al., 1993); *gyrA* using EfmE1679_2520 to determine amino acid changes of E87K/G or S83R/Y/I (Leavis et al., 2006b); *parC* using EfmE1679_0369 to determine amino acid changes of E86K or S82R/I (Leavis et al., 2006b); and *pbp5* (GenBank accession no. ZP_00603984) to search for the low-affinity *pbp5* consensus sequence (Galloway-Pena et al., 2011; Rice et al., 2004). (Qin et al., 2012)

[3.3] CHAPTER 2 RESULTS

As discussed in section 3.1, fundamental differences at the core genome level have been suggested in previous publications (van Schaik et al., 2010) and this was supported by the finding in the previous chapter that strains separated clearly into two distinct groups based on four genes, with a 3% nucleotide difference for *gls20*, 5% for *pbp5*, 7% for *pbp2*, and 10% for *wlcA* between strains in the two groups. In this section, SNP, phylogenetic, and 16S rRNA analyses showed differences between HA-clade strains and CA-clade strains throughout the core genome. These core genomic differences were then used to make molecular clock estimations to show that divergence of these two clades was a distant evolutionary event. Additionally, data are shown from a second study in which whole-genomic (using the TX16 closed genome and 20 draft genomes) analyses confirmed that there are differences between the HA- clade and CA- clade throughout the core genome. Using these genome sequences, CA and HA-clade isolates were further analyzed for the presence of unique IS elements, transposons, antibiotic resistance determinants, CRISPRS and putative virulence determinants.

[3.3.1] Analysis of 100 genes from 21 *E. faecium* genomes illustrates the differences in core genome components of two clades.

> A total of 1608 orthologs (defined as >80% identity and aligning to >60% match length) were common to all 21 genomes; 638 of these had the same DNA length in the available sequences from all the genomes analyzed. The 100 gene sequences chosen (Table 10 and Figure 7) had $\geq$ 95% identity and were concatenated into a 106,818bp nucleotide sequence for each of the strains (information on genomes listed in Table 9). The comparisons between all the strains were determined and are shown in Table 11. (Galloway-Pena et al., 2012)

TABLE 11. Percent identity and divergence score matrix of the 100 concatenated gene nucleotide sequence. (on next page) Table from (Galloway-Pena et al., 2012).

Percent identity

Divergence

	1141733[a]	Com12	Com15	E980	TX1330	1231408[b]	1230933[c]	1231410	1231501	1231502	C68	D344	TX16	E1039	E1071	E1162	E1636	E1679	TX82	TX0133	U0317
1141733	***	99.8	98.9	98.8	99.8	96.8	95.9	96	96.2	96	95.8	95.9	95.9	96.1	96	95.9	96	95.9	95.9	95.8	96
Com12	0.2	***	98.9	98.9	100	96.7	96	96	96.2	96	95.8	95.9	96	96.2	96	95.9	96	96	95.9	95.9	96
Com15	1.1	1.1	***	99.2	98.9	96.6	95.9	96	96.2	96.1	95.8	95.9	95.9	96.2	96.1	96	96.1	96	96	95.9	96.1
E980	1.2	1.1	0.8	***	98.9	96.6	96.2	96.3	96.5	96.3	96	96.2	96.2	96.5	96.3	96.2	96.3	96.2	96.2	96.2	96.3
TX1330	0.2	0	1.1	1.1	***	96.7	96	96	96.2	96	95.8	95.9	96	96.2	96	95.9	96	96	95.9	95.9	96
1231408	3.3	3.4	3.5	3.5	3.4	***	98.9	98.8	98.4	98.9	98.7	98.8	98.9	98.5	98.8	98.9	98.5	98.7	98.9	98.9	98.9
1230933	4.2	4.2	4.2	3.9	4.2	1.1	***	99.9	99.2	99.8	99.6	99.7	100	99.2	99.6	99.8	99.4	99.6	99.8	99.8	99.8
1231410	4.1	4.1	4.1	3.9	4.1	1.2	0.1	***	99.2	99.8	99.7	99.6	99.9	99.2	99.6	99.9	99.3	99.5	99.9	99.8	99.8
1231501	3.9	3.9	3.9	3.6	3.9	1.6	0.8	0.9	***	99.1	99	99.3	99.2	99.2	99.1	99.2	99.4	99.2	99.2	99.3	99.1
1231502	4.1	4.1	4.1	3.8	4.1	1.1	0.2	0.2	0.9	***	99.5	99.5	99.8	99.2	99.7	99.7	99.2	99.5	99.7	99.7	100
C68	4.4	4.4	4.4	4.1	4.4	1.3	0.4	0.3	1	0.5	***	99.5	99.6	99	99.5	99.8	99.2	99.4	99.8	99.7	99.5
D344SRF	4.3	4.2	4.2	3.9	4.2	1.2	0.3	0.4	0.7	0.5	0.5	***	99.7	99.2	99.5	99.7	99.7	99.7	99.7	99.8	99.5
TX16	4.2	4.2	4.2	3.9	4.2	1.1	0	0.1	0.8	0.2	0.4	0.3	***	99.2	99.6	99.8	99.4	99.5	99.8	99.8	99.8
E1039	4	4	3.9	3.7	4	1.5	0.8	0.8	0.8	0.8	1	0.8	0.8	***	99.2	99.2	99.5	99.2	99.2	99.2	99.2
E1071	4.1	4.1	4.1	3.8	4.1	1.2	0.4	0.4	0.9	0.3	0.5	0.5	0.4	0.8	***	99.7	99.3	99.6	99.7	99.6	99.7
E1162	4.2	4.2	4.2	3.9	4.2	1.1	0.2	0.1	0.8	0.3	0.2	0.3	0.2	0.8	0.3	***	99.4	99.6	100	99.9	99.7
E1636	4.1	4.1	4.1	3.8	4.1	1.5	0.6	0.7	0.6	0.8	0.8	0.3	0.6	0.5	0.7	0.6	***	99.5	99.4	99.5	99.2
E1679	4.2	4.2	4.2	3.9	4.2	1.3	0.4	0.5	0.8	0.5	0.6	0.3	0.5	0.8	0.4	0.4	0.5	***	99.6	99.7	99.5
TX82	4.2	4.2	4.2	3.9	4.2	1.1	0.2	0.1	0.8	0.3	0.2	0.3	0.2	0.8	0.3	0	0.6	0.4	***	99.9	99.7
TX0133A	4.3	4.3	4.3	4	4.3	1.1	0.2	0.2	0.7	0.3	0.3	0.2	0.2	0.8	0.4	0.1	0.5	0.3	0.1	***	99.7
U0317	4.1	4.1	4.1	3.8	4.1	1.2	0.2	0.2	0.9	0	0.5	0.5	0.2	0.8	0.3	0.3	0.8	0.5	0.3	0.3	***

TABLE 11 (continued).

[a] The numbers in yellow are the percent identity and divergence scores of the CA strains

[b] The numbers in blue are the percent identity and divergence scores of the hybrid strain

[c] The numbers in green are the percent identity and divergence scores of the HA strains

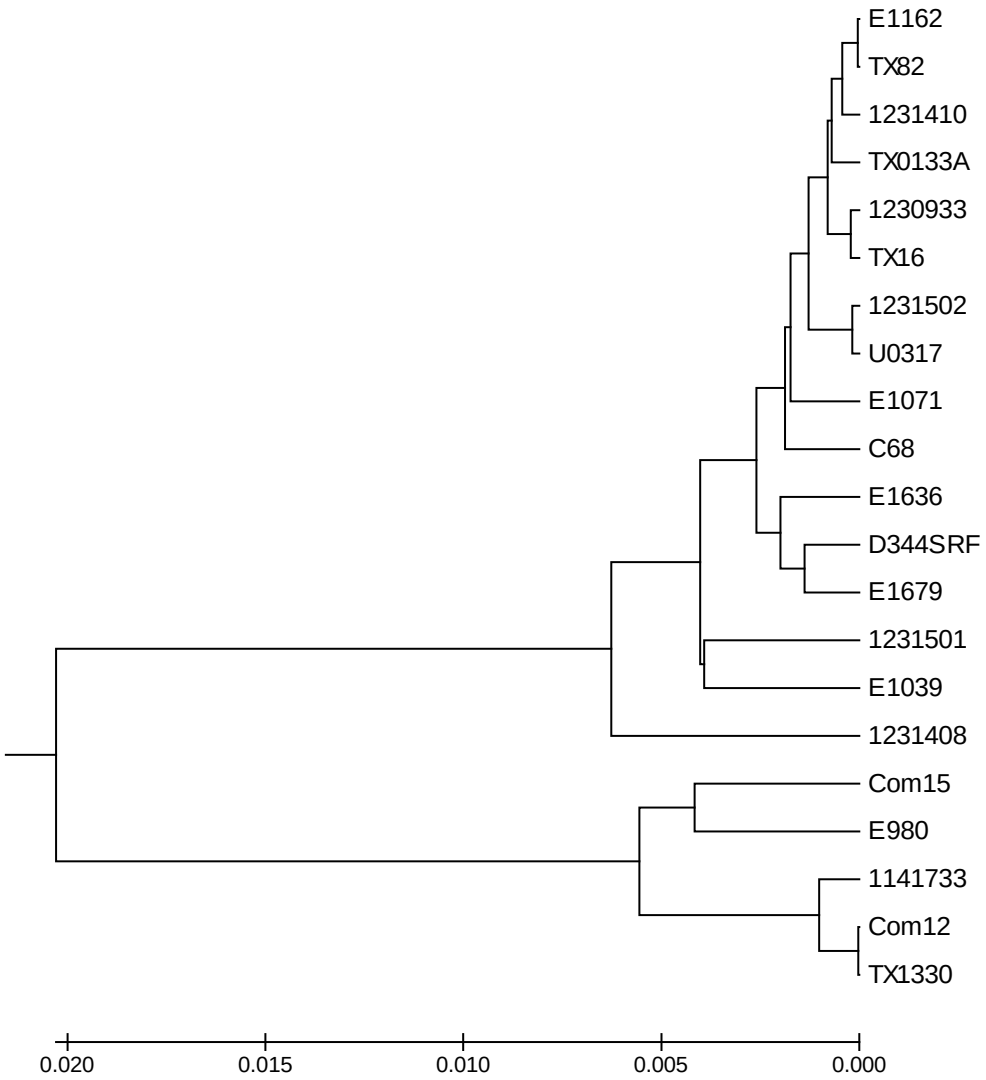


FIGURE 8. UPGMA phylogenetic tree of the concatenated 100 gene sequence (106,818 bp). The results are based on a pair-wise analysis of the concatenated 100 gene nucleotide sequence for each of the 21 *E. faecium* strains using the Poisson correction method in MEGA4.0.2 software. An UPGMA tree was constructed depicting the evolutionary distance between these sequences. The tree is drawn to scale with the branch lengths representing the evolutionary distances, the scale can be seen at the bottom of the tree. Figure from (Galloway-Pena et al., 2012)

The nucleotide percent identities among the strains within the CA clade (in yellow) ranged from 98.9%-100%, and the divergence scores from 0-1.2; the nucleotide percent identities between the strains in the HA clade (in green) range from 99-100% and the divergence scores from 0-1 (Table 11). However, when the two clades were compared to each other, the percent identities ranged from 95.8-96.5% and the divergence scores from 3.6-4.3 (Table 11). Therefore, 1141733, Com12, Com15, E980, and TX1330 are more closely related to each other than to the other 16 *E. faecium* strains and in turn, these 16 strains were found to be more related to each other than to the above 5 strains, reinforcing the grouping into two distinct clades.

The phylogenetic tree constructed using the concatenated sequences is shown in Figure 8. The tree separates the genomes into two distinct groups with one group consisting primarily of the strains from healthy volunteer fecal samples (Com12, Com15, TX1330, and E980), while most of the strains from clinical samples were in the other branch. Similar to previous terminology (Leavis et al., 2007; van Schaik et al., 2010), we designated these two branches the community-associated (CA) clade and the hospital-associated (HA) clade, respectively. There were, however, two exceptions. Strain 1141733 (which is a blood culture from a hospitalized patient) grouped in the branch with the commensal, fecal strains, and E1039, which is a healthy volunteer fecal sample, grouped with the clinical strains.

One *E. faecium* strain, 1231408 (in blue), isolated from the blood of a hospitalized patient, was found to have a "hybrid" genome sequence of both groups. The sequence between ORF No. 10017 and 10683 of 1231408 was found to be more similar to the sequences of strains in the CA clade whereas the rest of the 1231408 sequence was more similar to *E. faecium* strains in the HA clade. The overall percent identities of 1231408 ranged from 98.4-98.9% against HA clade strains and 96.7-96.8% against CA clade strains with divergence scores ranging from 1.1-1.6 against HA clade strains, and 3.3-3.5 against CA clade strains (Table 11).

The sequences of the individual genes from each of the 21 genomes were also compared and UPGMA trees constructed (data not shown) to create 100 individual trees; 92 of the 100 genes split the genomes into two distinct groups, while the other 8 genes were either 100% conserved or the sequences were not distinct enough to give two clear grouping. For each genome, we counted the genes that segmented into the HA clade or the CA clade (Table 12). The CA clade strains ranged from strain 1141733 with 92 of 92 genes in the CA clade to E980 with 84/92 genes in the CA clade and 8/92 genes within the HA clade. The HA clade strains ranged from D344SRF, TX82, TX0133, E1162, and C68, all with 92/92 genes grouping with the HA clade strains, to E1039 with 86/92 genes grouping with the HA clade and 6/92 genes in the CA clade. The hybrid strain 1231408 had 66/92 genes group with the HA clade and the other 26 with the CA clade, consistent with the percent identity and divergence scores above.

The encoded proteins of the 100 genes were also concatenated, resulting in a 35,606 amino acid sequence. The percent identities and divergence scores were tabulated as well as a UPGMA phylogenetic tree constructed for the concatenated protein sequence (Table 13 and Figure 9, respectively). As expected, the results were similar to the nucleotide results, grouping the strains into two distinct clades. (Galloway-Pena et al., 2012)

TABLE 12. Clade analysis of 91 individual genes.

Community Clade	Hospital Clade	Hybrid Clade
1141733 (92 + 0)[a]	C68 (0 + 92)	1231408 (26 + 66)
Com12 (91 + 1)	D344SRF (0 + 92)	
TX1330 (91 + 1)	E1162 (0 + 92)	
Com15 (90 + 2)	TX82 (0 + 92)	
E980 (84 + 8)	TX0133 (0 + 92)	
	TX16 (1 + 91)	
	1230933 (1 + 91)	
	1231410 (1 + 91)	
	E1071 (2 + 90)	
	E1636 (2 + 90)	
	E1679 (4 + 88)	
	U0317 (4 + 88)	
	1231502 (4 + 88)	
	E1039 (5 + 87)	
	1231501 (5 + 87)	

[a] Numbers in parentheses are given as number of genes which grouped with the community-associated clade plus the number of genes that grouped with the hospital-associated clade.

Table from (Galloway-Pena et al., 2012).

TABLE 13. Percent identity and divergence score matrix of the concatenated amino acid sequence. (on next page) Table from (Galloway-Pena et al., 2012).

Percent identity

Divergence

	1141733	Com12	Com15	E980	TX1330	1231408	1230933	1231410	1231501	1231502	C68	D344	TX16	E1039	E1071	E1162	E1636	E1679	TX82	TX0133	U0317
1141733	***	99.8	99.6	99.5	99.8	99.1	98.7	98.7	98.8	98.7	98.7	98.7	98.7	98.8	98.7	98.7	98.7	98.7	98.7	98.7	98.7
Com12	0.2	***	99.5	99.6	100	99.1	98.7	98.7	98.8	98.7	98.7	98.7	98.7	98.8	98.7	98.7	98.8	98.7	98.7	98.7	98.8
Com15	0.4	0.5	***	99.6	99.5	99	98.7	98.7	98.8	98.7	98.7	98.7	98.7	98.8	98.7	98.7	98.7	98.7	98.7	98.7	98.7
E980	0.5	0.4	0.4	***	99.6	99	98.8	98.8	98.9	98.8	98.8	98.8	98.8	98.9	98.8	98.8	98.8	98.8	98.8	98.8	98.8
TX1330	0.2	0	0.5	0.4	***	99.1	98.7	98.7	98.8	98.7	98.7	98.7	98.7	98.8	98.7	98.7	98.8	98.7	98.7	98.7	98.8
1231408	0.9	0.9	1	1	0.9	***	99.5	99.5	99.3	99.5	99.5	99.4	99.5	99.3	99.5	99.5	99.3	99.4	99.5	99.5	99.5
1230933	1.3	1.3	1.3	1.2	1.3	0.5	***	100	99.7	99.9	99.9	99.8	99.9	99.7	99.9	99.9	99.7	99.8	99.9	99.9	99.9
1231410	1.3	1.3	1.3	1.2	1.3	0.5	0	***	99.7	99.9	99.9	99.8	99.9	99.7	99.9	100	99.7	99.8	100	99.9	99.9
1231501	1.2	1.2	1.2	1.1	1.2	0.7	0.3	0.3	***	99.7	99.7	99.6	99.7	99.7	99.6	99.7	99.7	99.6	99.7	99.7	99.7
1231502	1.3	1.3	1.3	1.2	1.3	0.5	0.1	0.1	0.3	***	99.9	99.8	99.9	99.7	99.9	99.9	99.7	99.8	99.9	99.9	100
C68	1.3	1.3	1.3	1.2	1.3	0.5	0.1	0.1	0.3	0.1	***	99.8	99.9	99.6	99.8	99.9	99.7	99.8	99.9	99.9	99.9
D344SRF	1.3	1.3	1.3	1.2	1.3	0.6	0.2	0.2	0.4	0.2	0.2	***	99.8	99.6	99.8	99.8	99.9	99.8	99.8	99.8	99.8
TX16	1.3	1.3	1.3	1.2	1.3	0.5	0.1	0.1	0.3	0.1	0.1	0.2	***	99.6	99.8	99.9	99.7	99.8	99.9	99.9	99.9
E1039	1.3	1.2	1.3	1.1	1.2	0.7	0.3	0.3	0.3	0.3	0.4	0.4	0.4	***	99.7	99.7	99.7	99.6	99.7	99.7	99.7
E1071	1.3	1.3	1.3	1.2	1.3	0.5	0.1	0.1	0.4	0.1	0.2	0.2	0.2	0.3	***	99.9	99.7	99.8	99.9	99.9	99.9
E1162	1.3	1.3	1.3	1.2	1.3	0.5	0.1	0	0.3	0.1	0.1	0.2	0.1	0.3	0.1	***	99.7	99.8	100	100	99.9
E1636	1.3	1.3	1.3	1.2	1.3	0.7	0.3	0.3	0.3	0.3	0.3	0.1	0.3	0.3	0.3	0.3	***	99.8	99.7	99.7	99.7
E1679	1.3	1.3	1.3	1.2	1.3	0.6	0.2	0.2	0.4	0.2	0.2	0.2	0.2	0.4	0.2	0.2	0.2	***	99.8	99.8	99.8
TX82	1.3	1.3	1.3	1.2	1.3	0.5	0.1	0	0.3	0.1	0.1	0.2	0.1	0.3	0.1	0	0.3	0.2	***	99.9	99.9
TX0133A	1.3	1.3	1.3	1.2	1.3	0.5	0.1	0.1	0.3	0.1	0.1	0.2	0.1	0.3	0.1	0	0.3	0.2	0.1	***	99.9
U0317	1.3	1.3	1.3	1.2	1.3	0.5	0.1	0.1	0.3	0	0.1	0.2	0.1	0.3	0.1	0.1	0.3	0.2	0.1	0.1	***

TABLE 13. (continued)

[a] The numbers in yellow are the percent identity and divergence scores of the CA strains

[b] The numbers in blue are the percent identity and divergence scores of the hybrid strain

[c] The numbers in green are the percent identity and divergence scores of the HA strains

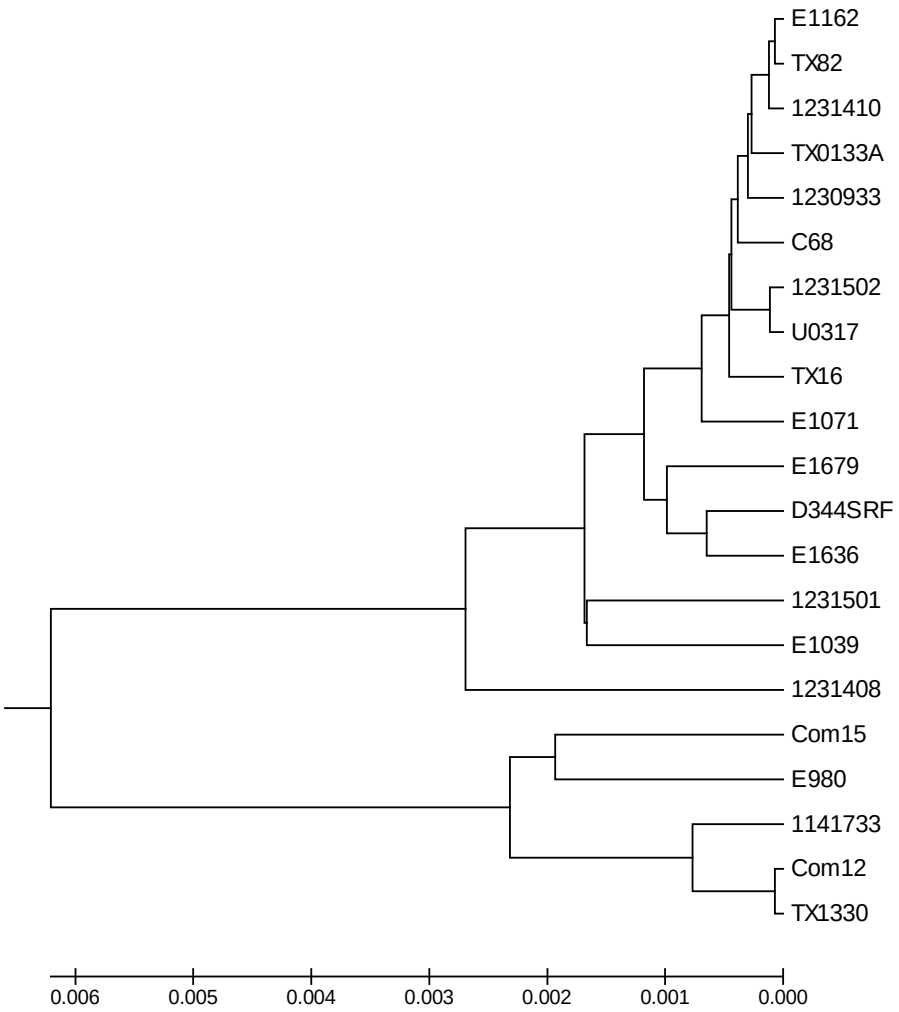


FIGURE 9. UPGMA phylogenetic tree of the concatenated amino acid sequence (35,606bp). The results are based on a pair-wise analysis of the concatenated 100 protein amino acid sequence for each of the 21 *E. faecium* strains using the Poisson correction method in MEGA4.0.2 software. An UPGMA tree was constructed depicting the evolutionary distance between these sequences. The tree is drawn to scale with the branch lengths representing the evolutionary distances, the scale can be seen at the bottom of the tree. Figure from (Galloway-Pena et al., 2012).

[3.3.2] SNP analysis emphasizes clade-specific differences.

5,932 SNP sites (5.6% of the total sequence) were identified in the 106,818 bp aligned sequences. Among these, 929 SNPs were strain specific SNPs (identified in only one strain) and the remaining 5,003 SNPs were identified in two or more *E. faecium* strains. Each SNP was concatenated into a continuous sequence; the percent identities and divergent scores for this analysis can be seen in Table 14 and the UPGMA tree seen in Figure 10. The percent identities of the concatenated SNP sequences among the CA clade strains ranged from 79-100% and the divergence scores from 0-26. Within the HA clade, the percent identities of the SNP sequences ranged from 82-100% and the divergence scores from 0-21. When the two clades were compared by pair-wise analysis of each strain to one another, the percent identities ranged from 24-37% and the divergence score was 350. For the hybrid strain 1231408, the identities ranged from 38-42% and the divergence scores 23-39 with CA clade strains and 71-81% and 350 with HA clade strains.

Of the 5,932 SNPs, 5,147 (86.8%) were synonymous SNPs, whereas 785 (13.2%) were non-synonymous and found in 87 out of the 100 genes. Using the criteria described in the methods, 479 of the 785 non-synonymous SNPs were considered clade specific and were found in 72 of the 100 genes (Table 10). When searching through the KEGG map, the clade specific SNPs were found in 33 different metabolic pathways and 5 different genetic/ environmental/cellular processes (Table 10). (Galloway-Pena et al., 2012)

TABLE 14. Percent identity and divergence score matrix of the concatenated SNP sequence. (on next page) Table from (Galloway-Pena et al., 2012).

Percent identity

Divergence

	1141733	Com12	Com15	E980	TX1330	1231408	1230933	1231410	1231501	1231502	C68	D344	TX16	E1039	E1071	E1162	E1636	E1679	TX82	TX0133	U0317
1141733	***	96	81	79	96	42	27	28	32	28	24	26	27	30	28	27	28	27	26	25	28
Com12	4	***	80	80	100	41	27	28	32	29	24	26	27	31	29	27	29	27	27	26	29
Com15	23	24	***	85	80	38	27	28	32	29	24	26	27	32	29	27	29	28	27	26	29
E980	26	24	17	***	80	39	32	33	37	34	29	31	32	36	33	31	34	32	31	31	34
TX1330	4	0	24	24	***	41	27	28	32	29	24	26	27	31	29	27	29	27	27	26	29
1231408	350	350	350	350	350	***	80	79	71	80	77	78	80	74	78	80	73	76	80	81	79
1230933	350	350	350	350	350	24	***	98	86	97	93	94	99	86	92	97	88	92	97	97	97
1231410	350	350	350	350	350	26	2	***	85	96	95	93	98	85	94	99	87	91	98	97	96
1231501	350	350	350	350	350	39	16	18	***	84	83	87	86	86	84	86	90	85	86	87	84
1231502	350	350	350	350	350	25	3	4	19	***	91	92	97	86	94	95	86	91	94	95	99
C68	350	350	350	350	350	28	8	5	20	10	***	91	93	82	91	96	85	89	96	94	90
D344SRF	350	350	350	350	350	27	7	8	15	9	10	***	94	86	91	94	94	95	94	96	91
TX16	350	350	350	350	350	25	1	2	16	3	7	6	***	86	92	97	88	92	97	97	97
E1039	350	350	350	350	350	35	16	17	16	16	21	16	16	***	86	85	91	85	85	86	86
E1071	350	350	350	350	350	28	8	7	19	6	10	10	8	16	***	95	87	92	95	93	95
E1162	350	350	350	350	350	24	4	2	16	6	4	6	3	17	5	***	89	92	100	98	94
E1636	350	350	350	350	350	36	13	14	12	15	17	6	13	10	14	12	***	92	89	90	86
E1679	350	350	350	350	350	30	8	10	17	10	13	5	9	17	9	8	9	***	92	94	91
TX82	350	350	350	350	350	24	4	2	16	6	4	6	3	17	5	0	13	9	***	98	94
TX0133A	350	350	350	350	350	23	3	4	15	5	6	5	3	16	7	2	11	7	2	***	95
U0317	350	350	350	350	350	25	3	5	19	1	10	9	4	16	6	6	15	10	6	6	***

TABLE 14. (continued)

[a] The numbers in yellow are the percent identity and divergence scores of the CA strains

[b] The numbers in blue are the percent identity and divergence scores of the hybrid strain

[c] The numbers in green are the percent identity and divergence scores of the HA strains

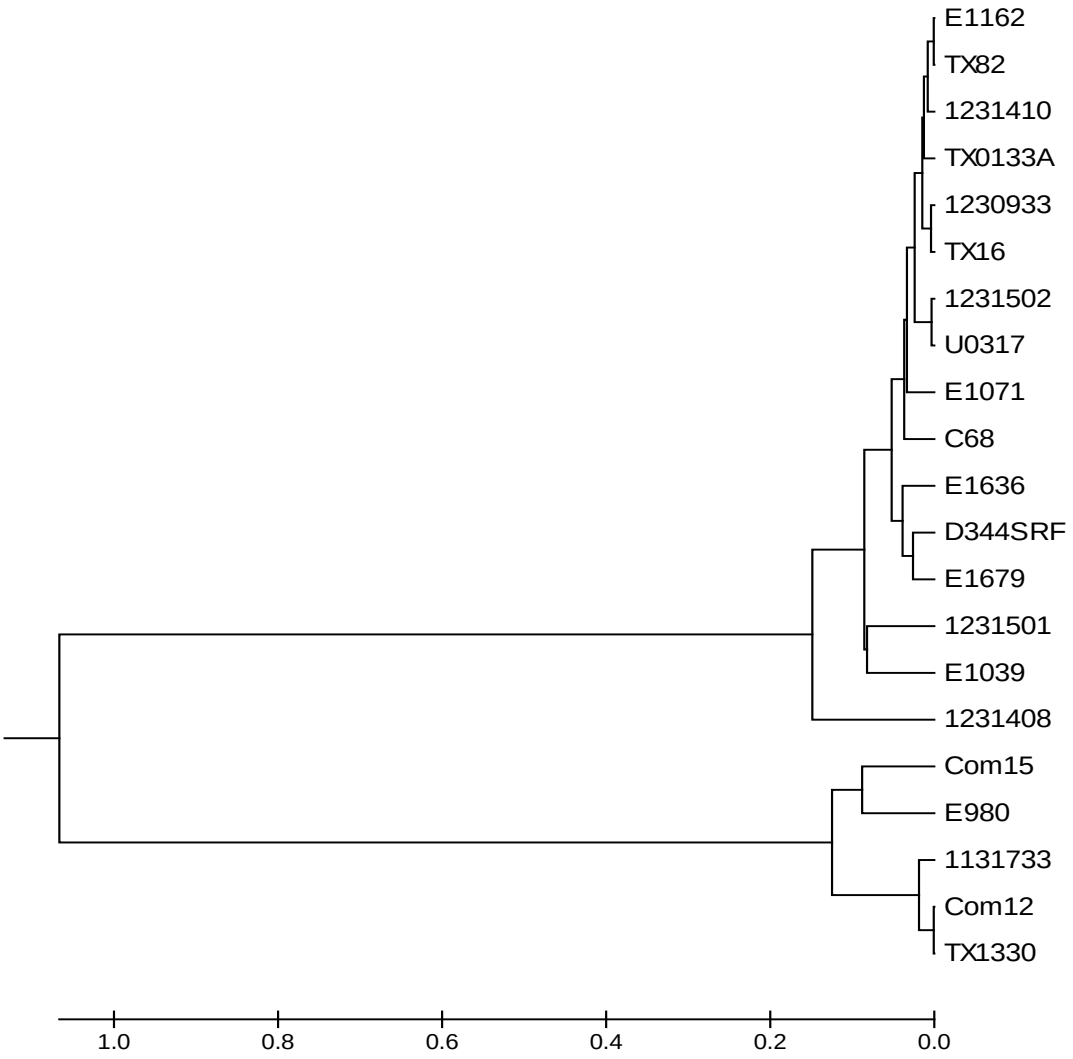


FIGURE 10. UPGMA phylogenetic tree of the concatenated SNP sequence (5,392 bp). The results are based on a pair-wise analysis of the concatenated SNP nucleotide sequence for each of the 21 *E. faecium* strains using the Poisson correction method in MEGA4.0.2 software. An UPGMA tree was constructed depicting the evolutionary distance between these sequences. The tree is drawn to scale with the branch lengths representing the evolutionary distances, the scale can be seen at the bottom of the tree. Figure from (Galloway-Pena et al., 2012).

[3.3.3] Molecular clock determination using sSNP analysis shows an evolutionarily distant separation between the CA and HA clades.

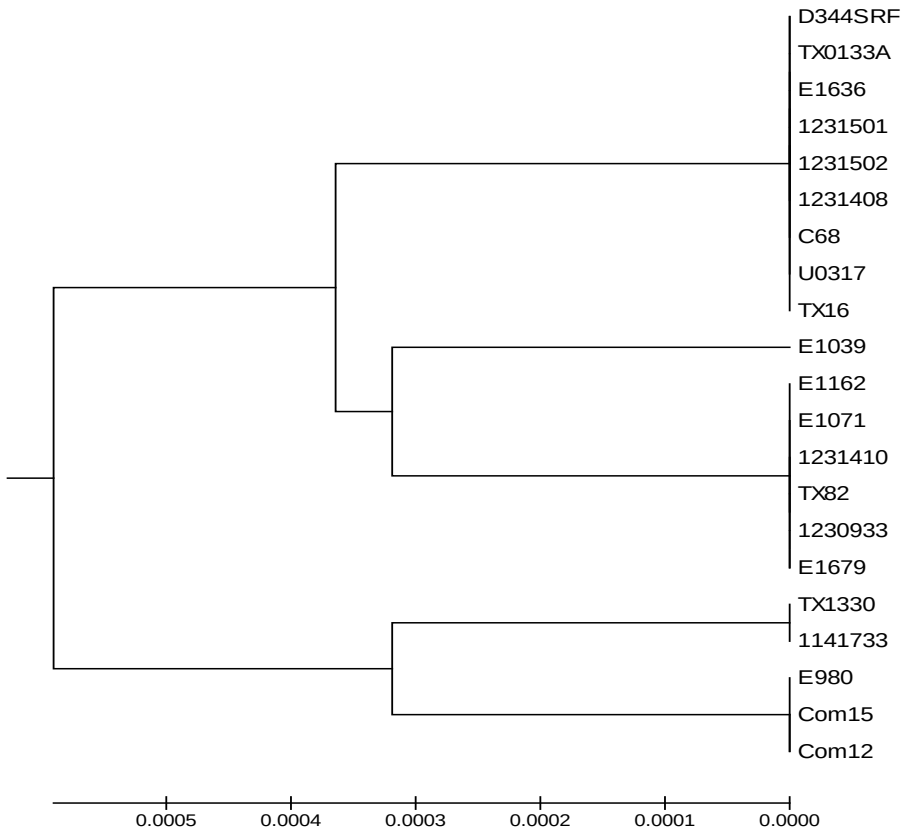
We estimated the time of divergence using sSNP (Foster et al., 2009; Lenski et al., 2003; Pearson et al., 2009; Van Ert et al., 2007). The median number of sSNP in the CA clade was 979 (ranging from 1-1077), the median number of sSNPs in the HA clade was 401 (ranging from 2-916), and the median number of sSNPs between the two clades was 3812 (ranging from 3368-4065). These numbers were used to calculate the molecular clock and the results are shown in Table 15. Using the *E. coli* synonymous mutation rate and a range of generation times of 100-300 per year, the molecular clock determination estimated that strains within the CA clade diverged from each other on the average of ~300,000 to 900,000 years ago, whereas strains within the HA clade diverged on the average of ~100,000-300,000 years ago. When we used the same formula to calculate the divergence between the two clades, it is estimated that the two clades diverged between one million and three million years ago. Even with the faster mutational rate previously used with *B. anthracis*, the time of divergence was estimated at 300,000 years between the two clades (Table 15). (Galloway-Pena et al., 2012)

TABLE 15. Molecular clock/time of divergence estimates based on sSNP analysis within and between clades.

Based on the mutation rate of *Escherichia coli* (1.4×10^{-10})			
Generations/yr	**Hospital Clade**	**Community Clade**	**Two clades**
100	380,535	929,036	3,617,453
200	190,267	464,518	1,808,726
300	126,845	309,679	1,205,818
Based on the mutation rate of *Bacillus anthracis* (5.2×10^{-10})			
Generations/yr	**Hospital Clade**	**Community Clade**	**Two clades**
100	102,452	250,125	973,930
200	51,226	125,063	486,965
300	34,151	83,375	324,643

[3.3.4] 16S rRNA and ribosomal protein analyses show clade-specific SNPs and a similar time of divergence to the SNP analysis.

One specific SNP at base-pair 61 out of 1569 differed between the strains following the CA and HA clade grouping as seen in the other analyses (Figure 11). CA clade strains had a thymine at this position, while HA clade strains had an adenine. Two other SNPs occurred at positions 103 and 193 in some of the strains (Figure 11). Ribosomal protein L32 gene (*rpmF*) (Figure 12), as well as ribosomal proteins L35, L21, S18 and S16 showed that all CA clade strains differed from the HA clade strains by one to two nucleotides and there were no differences within a given clade. These results corroborated the molecular clock determination by SNP analysis. Using estimates of approximately 1% change every 25 million years in the 16S rRNA (Achtman et al., 2004; Ochman et al., 1999; Ochman and Wilson, 1987), a one to two nucleotide change out of 1569 bp (0.06-0.1%) would approximate the time of divergence of the two clades as somewhere between 1.5 million to 2.5 million years ago. (Galloway-Pena et al., 2012)



16S rRNA position		
61	103	193
A	G	G
A	A	A
A	G	A
T	A	G
T	G	G

FIGURE 11. An UPGMA tree representing the evolutionary relationship of the strains using the 16S rRNA nucleotide sequence for each of the 21 *E. faecium* strains. For all phylogenetic trees, the evolutionary distances were calculated using the Poisson correction method and UPMGA trees constructed using MEGA4.0.2 software. The trees are drawn to scale with the branch

lengths representing the evolutionary distances, the scale of each tree can be seen at the bottom of each respective tree. The table next to the tree indicates the changes in the 16S rRNA gene sequence at each nucleotide position for each branch of the tree. Figure from (Galloway-Pena et al., 2012).

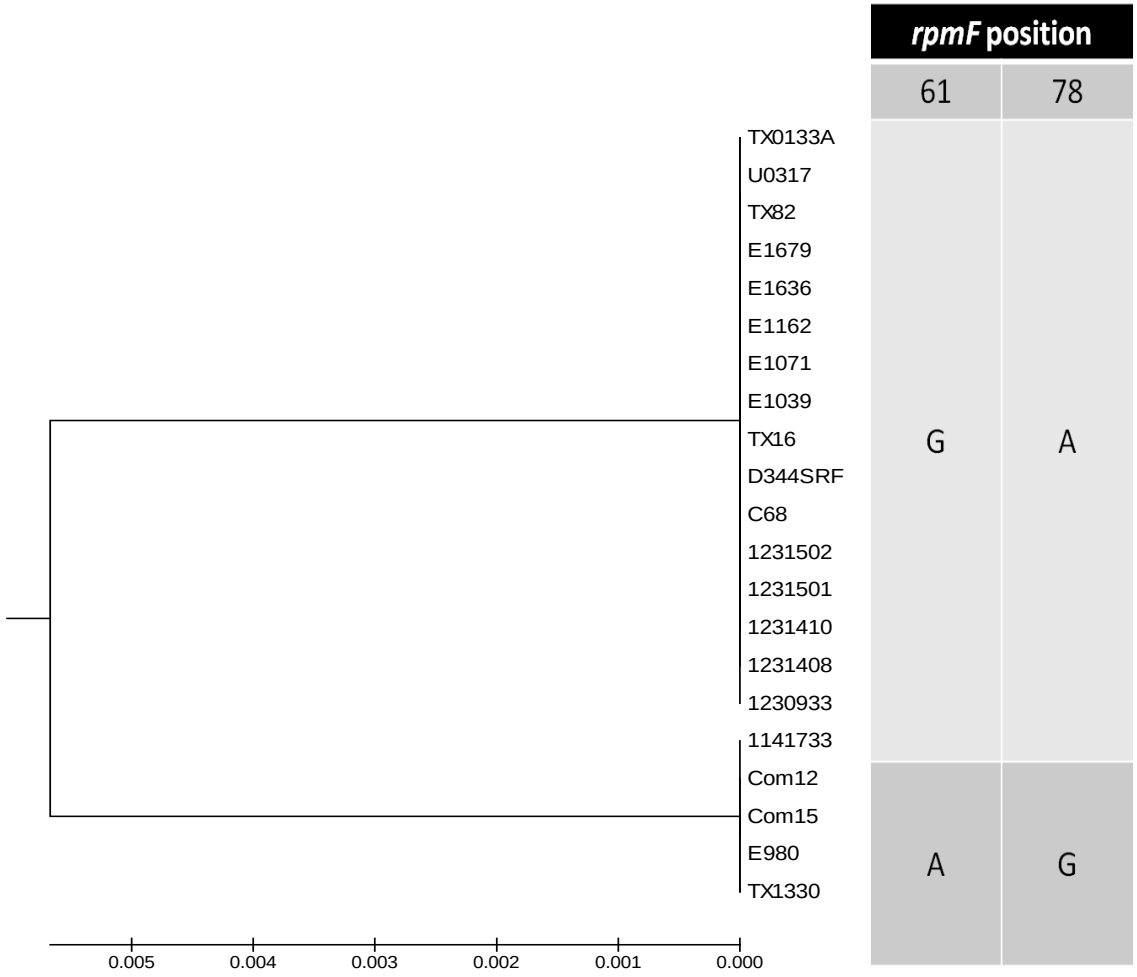


FIGURE 12. An UPGMA tree representing the evolutionary relationship of the strains using the *rpmF* nucleotide sequence for each of the 21 *E. faecium* strains. For all phylogenetic trees, the evolutionary distances were calculated using the Poisson correction method and UPMGA trees constructed using MEGA4.0.2 software. The trees are drawn to scale with the branch lengths representing

the evolutionary distances, the scale of each tree can be seen at the bottom of each respective tree. The table next to the tree indicates the changes in the *rpmF* sequence at each nucleotide position for each branch of the tree. Figure from (Galloway-Pena et al., 2012).

[3.3.5] Phylogenetic, multi-locus sequence typing (MLST) and gene content similarity analysis.

Analysis of the 21 *E. faecium* genomes (Table 9) showed that the isolates separate into two clades, one branch consisting mostly of CA isolates, with most HA isolates found in the other, as was noted in our previous study (Galloway-Pena et al., 2012) (Figure 13A and B). When analyzing the phylogenetic distances among these 21 isolates using 628 single-copy ortholog genes of the same length (Figure 13A), similar clade patterns were observed for the *E. faecium* strains as seen in the 100 core gene analysis by Galloway-Pena et.al (Galloway-Pena et al., 2012). All isolates predicted to be part of the CC17 genogroup (Leavis et al., 2006a; Top et al., 2008; Willems et al., 2011) cluster more closely together and branched more distantly than other HA-clade isolates (Figure 13A). The dendogram construction from the gene content dissimilarity represented by Jaccard distance (Figure 13B) also showed most hospital-isolated strains cluster together except hospital- isolated strain 1,141,733 which was shown genetically to belong to the CA clade. In addition, although E1039 is a community- isolated fecal strain, it is genetically closer to the HA strains. The phylogenetic and gene content dissimilarity analysis results all support the existence of two very distinct clades of *E. faecium*, which has been previously described using pyrosequencing, microarray, and the concatenation of a 100 core genes, estimated to have diverged anywhere from 300,000 to 3 million years ago (Galloway-Pena et al., 2012; Leavis et al., 2007; van Schaik et al., 2010). (Qin et al., 2012)

[3.3.6] Mobile genetic elements.

IS elements and transposases were found more frequently in HA strains than in CA strains. Previously, IS*16* was suggested as a molecular screening marker to predict *E. faecium* pathogenicity because of its presence in clinical *E. faecium* isolates (Leavis et al., 2007; Werner et al., 2011). We performed a BLAST search of the 22 *E. faecium* genomes to identify the IS/transposase elements showing the same presence or absence patterns of IS16 (HMPREF0351_11812, _11855, _12352, and _12809) (Table 16). Many IS/transposase elements were found to have the same pattern of presence/absence in different strains as IS16; including ISEnfa3 (IS3/IS911 transposase: HMPREF0351_10172, _10364, _11866, and _11868), IS116/IS110/IS902 family transposases (HMPREF0351_11035, _11528, _12768, and _13088), IS66 transposases (HMPREF0351_10928, _11787, _11933, _12004, _12887, and_12948), and transposases (HMPREF0351_10878, _10880, _10927, _11934, and _12005). Therefore, all these IS elements and transposases (in addition to IS*16*) have potential as molecular markers to identify clinical *E. faecium*. However, these IS elements and transposases are not found in all HA-clade strains as 1,231,501; E1039; and E1071 do not have these IS elements and transposases, although they are present in all of the isolates considered to be part of the CC17 genogroup (Figure 13A). (Qin et al., 2012)

TABLE 16. Hospital-associated clade unique mobile elements (IS elements/transposons)

TX16 gene ID	Present in numbers of isolates in the HA clade	Start	End	Strand	Predicted Gene Product	Gene product name
HMPREF0351_10172	13/16	169442	169978	-	.	IS3/IS911 family transposase
HMPREF0351_10314	10/16	293413	294708	+	.	IS204/IS1001/IS1096/IS1165 family transposase
HMPREF0351_10364	13/16	346595	347131	-	.	transposase IS3/IS911
HMPREF0351_10606	10/16	604854	606317	-	.	transposase
HMPREF0351_10878	12/16	857236	857526	+	.	IS3/IS911 transposase
HMPREF0351_10880	13/16	858483	858977	-	.	transposase
HMPREF0351_10925	12/16	902007	902753	-	.	transposase
HMPREF0351_10927	13/16	903083	903448	+	.	transposon protein
HMPREF0351_10928	13/16	903537	905084	+	.	transposase IS66
HMPREF0351_10995	10/16	971959	973254	-	.	transposase
HMPREF0351_11035	13/16	1020193	1021443	-	.	IS116/IS110/IS902 family transposase
HMPREF0351_11079	10/16	1063888	1065183	+	.	transposase
HMPREF0351_11528	13/16	1500554	1501804	+	.	IS116/IS110/IS902 family transposase
HMPREF0351_11787	13/16	1754624	1756171	-	.	transposase IS66
HMPREF0351_11788	12/16	1756260	1756625	-	.	transposon protein

HMPREF0351_11803	12/16	1770952	1771242	-	.	IS3/IS911 family transposase
HMPREF0351_11806	9/16	1773745	1774866	-	.	transposase
HMPREF0351_11807	9/16	1775185	1776603	+	.	transposase
HMPREF0351_11809	9/16	1777456	1778493	-	.	conjugative transposon protein
HMPREF0351_11812	13/16	1781818	1783005	-	.	IS256 family transposase
HMPREF0351_11814	13/16	1785430	1785819	-	.	conjugative transposon protein
HMPREF0351_11826	13/16	1797041	1797415	-	.	conjugative transposon protein
HMPREF0351_11827	13/16	1797428	1797742	-	.	conjugative transposon protein
HMPREF0351_11846	7/16	1815331	1816500	+	.	transposase
HMPREF0351_11848	11/16	1816704	1817516	+	.	transposase
HMPREF0351_11855	13/16	1821214	1822401	-	.	transposase
HMPREF0351_11858	13/16	1826059	1826295	+	.	IS116/IS110/IS902 family transposase
HMPREF0351_11859	13/16	1826261	1826491	+	.	IS116/IS110/IS902 family transposase
HMPREF0351_11860	13/16	1826473	1826772	+	.	IS116/IS110/IS902 family transposase
HMPREF0351_11861	13/16	1826769	1827311	+	.	IS116/IS110/IS902 family transposase
HMPREF0351_11866	13/16	1829580	1830116	-	.	transposase IS3/IS911
HMPREF0351_11868	13/16	1831575	1832111	+	.	transposase IS3/IS911
HMPREF0351_11914	5/16	1872987	1873493	+	.	IS4 family transposase
HMPREF0351_11915	5/16	1873490	1874071	+	.	IS4 family transposase

HMPREF0351_11933	13/16	1885081	1886628	-	.	transposase IS66
HMPREF0351_11934	13/16	1886730	1887083	-	.	transposase
HMPREF0351_12004	13/16	1960073	1961620	-	.	IS66 transposase
HMPREF0351_12005	13/16	1961722	1962075	-	.	transposase
HMPREF0351_12352	13/16	2296940	2298127	-	.	transposase
HMPREF0351_12420	10/16	2359695	2360990	-	.	transposase

Table from (Qin et al., 2012)

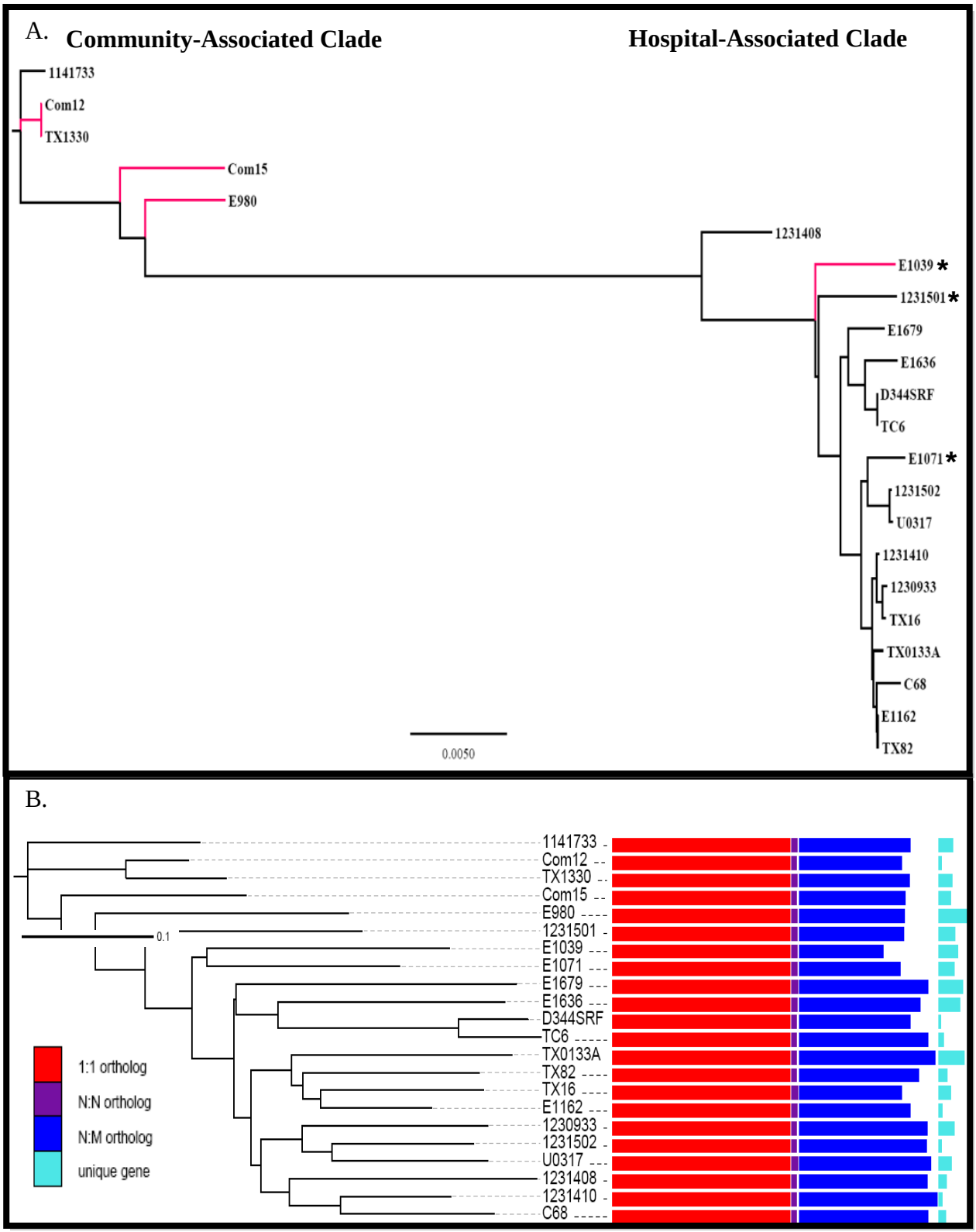

A.
Community-Associated Clade
Hospital-Associated Clade
1141733
Com12
TX1330
Com15
E980
1231408
E1039 *
1231501 *
E1679
E1636
D344SRF
TC6
E1071 *
1231502
U0317
1231410
1230933
TX16
TX0133A
C68
E1162
TX82
0.0050
B.
0.1
1141733
Com12
TX1330
Com15
E980
1231501
E1039
E1071
E1679
E1636
D344SRF
TC6
TX0133A
TX82
TX16
E1162
1230933
1231502
U0317
1231408
1231410
C68
1:1 ortholog
N:N ortholog
N:M ortholog
unique gene

FIGURE 13. *Enterococcus faecium* phylogenetics. 4A. A maximum-likelihood phylogenetic tree using 628 core genes. Distance bar indicates the sequence divergence. Strains isolated from the community are labeled with branches in red. An asterisk (*) indicates a strain within the HA clade lacking IS*16*. 4B. A hierarchical clustering using Jaccard distance of gene content by unweighted pair group method with arithmetic mean (UPGMA) (see Materials and Methods). The core, distributed and unique gene counts are also presented in the right panel. 1:1 ortholog, orthologs present with one copy in all strains; N:N ortholog, orthologs present with multiple copies in all strains; N:M ortholog, orthologs present in some strains. Figure from (Qin et al., 2012).

TABLE 17. Presence of genes encoding MSCRAMMs and pilins among 21 *E. faecium* genomes.

	***ebp* operon**			***fms11-fms19-fms16* cluster**			***fms21-fms20* cluster**	
Strain	***ebpA***	***ebpB***	***ebpC***	***fms11***	***fms19***	***fms16***	***fms21***	***fms20***
HA Clade								
TX16	*ebp*A var. 1[a]	*ebpB* var. 1	*ebpC* var. 1	*fms11* var. 1	*fms19* var. 1(p)[b]	*fms16* var. 1(p)	*fms21*	*fms20* var. 1
1,231,502	*ebp*A var. 1	*ebpB* var. 1	*ebpC* var. 1	*fms11* var. 1	*fms19* var. 1(p)	*fms16* var. 1(p)	*fms21*	*fms20* var. 1
1,231,410	*ebpA* var. 1 (p)	*ebpB* var. 1	*ebpC* var. 1	*fms11* var. 1	*fms19* var. 1(p)	*fms16* var. 1(p)	*fms21* (p)	"*fms20'* var. 2[c]
1,231,501	*ebp*A var. 1	*ebpB* var. 1	*ebpC* var. 1	*fms11* var. 1	*fms19* var. 1	*fms16* var. 2[d]	*fms21*	*fms20* var. 1
1,230,933	*ebp*A var. 1	*ebpB* var. 1	*ebpC* var. 1	*fms11* var. 1	*fms19* var. 1(p)	*fms16* var. 1(p)	*fms21* (p)	*fms20* var. 1
C68	*ebpA* var. 1 (p)	*ebpB* var. 1	*ebpC* var. 1	*fms11* var. 1	*fms19* var. 1(p)	*fms16* var. 1(p)	*fms21*	*fms20* var. 1
D344SRF	-	-	-	*fms11* var. 1	*fms19* var. 1(p)	*fms16* var. 1(p)	*fms21* (p) *fms21* (p)	*fms20* var. 1 "*fms20"* var. 2
U0137	*ebp*A var. 1	*ebpB* var. 1	*ebpC* var. 1	*fms11* var. 1(p)	*fms19* var. 1	*fms16* var. 1(p)	*fms21*	*fms20* var. 1
E1162	*ebp*A var. 1	*ebpB* var. 1	*ebpC* var. 1	*fms11* var. 1	*fms19* var. 1(p)	*fms16* var. 1(p)	*fms21*	*"fms20"* var. 2
E1636	-	-	-	-	-	-	*fms21* (p) *fms21* (p)	*fms20* var. 1 (p)
E1679	-	-	-	*fms11* var. 1	*fms19* var. 1(p)	*fms16* var. 1(p)	-	-
TX82	*ebp*A var. 1	*ebpB* var. 1	*ebpC* var. 1	*fms11* var. 1	*fms19* var. 1(p)	*fms16* var. 1(p)	-	-
TX0133A	*ebp*A var. 1	*ebpB* var. 1 (p)	*ebpC* var. 1	*fms11* var. 1	*fms19* var. 1(p)	*fms16* var. 1(p)	*fms21*	"*fms20"* var. 2
E1071	*ebp*A var. 1 (p)[e]	-	*ebpC* var. 1 (p)[e]	-	-	-	*fms21*	*fms20* var. 1
E1039	-	-	-	*fms11* var. 1(p)	-	-	*fms21*	*fms20* var. 1
1231408[f]	*ebp*A var. 1	*ebpB* var. 1	*ebpC* var. 1	*fms11* var. 1	*fms19* var. 1(p)	*fms16* var. 1(p)	*fms21*	*fms20* var. 1
CA Clade								
1,141,733	*ebp*A var. 2[g]	*ebpB* var. 2[h]	*ebpC* var. 2[i]	*fms11* var. 2[j]	*fms19* var. 2[k]	*fms16* var. 2	*fms21*	*fms20* var. 1
TX1330	*ebpA* var. 1 (p) *ebpA* var. 2 (p)	*ebpB* var. 1 *ebpB* var. 2 (p)	*ebpC* var. 1 *ebpC* var. 2	-	-	-	*fms21*	*fms20* var. 1
Com12	*ebpA* var. 1 (p) *ebpA* var. 2 (p)	*ebpB* var. 1 *ebpB* var. 2 (p)	*ebpC* var. 1 *ebpC* var. 2	-	-	-	*fms21*	*fms20* var. 1
Com15	*ebpA* var. 1 (p) *ebpA* var. 2 (p)	*ebpB* var. 1 *ebpB* var. 2	*ebpC* var. 1 (p) *ebpC* var. 2 (p)	*fms11* var. 2 (p)	*fms19* var. 2(p)	*fms16* var. 2(p)	*fms21*	*fms20* var. 1
E980	*ebp*A var. 2 (p)	*ebpB* var. 2	*ebpC* var. 2 (p)	*fms11* var. 2 (p)	*fms19* var. 2(p)	*fms16* var. 2	*fms21*	*fms20* var. 1

Strain	*fms14-fms17-fms13* cluster						
	fms14	***fms17***	***fms13***	***scm***	***fms18***	***fms15***	***acm***
HA Clade							
TX16	*fms14* var. 1	*fms17* var. 1	*fms13* var. 1	*scm* var. 1	*fms18* var. 1	*fms15* (p)	*acm*
1,231,502	*fms14* var. 1	*fms17* var. 1	*fms13* var. 1	*scm* var. 1 (p)	*fms18* var. 1 *fms18* var. 2[l]	*fms15* (p)	*acm*
1,231,410	*fms14* var. 1	*fms17* var. 1	*fms13* var. 1	*scm* var. 1 (p)	*fms18* var. 1 *fms18* var. 2	*fms15* (p)	*acm*
1,231,501	*fms14* var. 1	*fms17* var. 1	*fms13* var. 1	*scm* var. 1	-	*fms15* (p)	*acm*
1,230,933	*fms14* var. 1	*fms17* var. 1	*fms13* var. 1	*scm* var. 1 (p)	*fms18* var. 2	*fms15* (p)	*acm*
C68	*fms14* var. 1	*fms17* var. 1	*fms13* var. 1	*scm* var. 1 (p)	*fms18* var. 1 *fms18* var. 2	*fms15* (p)	*acm*
D344SRF	*fms14* var. 1 (p)	*fms17* var. 1 (p)	*fms13* var. 1 (p)	*scm* var. 1 (p)	*fms18* var. 1	*fms15*	*acm* (p)
U0137	*fms14* var. 1	*fms17* var. 1	*fms13* var. 1	*scm* var. 1 (p)	*fms18* var. 2	*fms15* (p)	*acm*
E1162	*fms14* var. 1	*fms17* var. 1	*fms13* var. 1	*scm* var. 1 (p)	*fms18* var. 1 *fms18* var. 2	*fms15*	*acm*
E1636	*fms14* var. 1	*fms17* var. 1 (p)	*fms13* var. 1 (p)	*scm* var. 1 (p)	*fms18* var. 2	*fms15*	*acm* (p)
E1679	*fms14* var. 1	*fms17* var. 1 (p)	*fms13* var. 1 (p)	*scm* var. 1	*fms18* var. 1 *fms18* var. 2	*fms15* (p)	*acm* (p)
TX82	*fms14* var. 1	*fms17* var. 1	*fms13* var. 1	*scm* var. 1 (p)	-	*fms15* (p)	*acm*
TX0133A	*fms14* var. 1	*fms17* var. 1 (p)	*fms13* var. 1 (p)	*scm* var. 1 (p)	*fms18* var. 1 *fms18* var. 2	*fms15* (p)	*acm*
E1071	*fms14* var. 1	*fms17* var. 1	*fms13* var. 1	*scm* var. 1 (p)	*fms18* var. 2	*fms15* (p)	*acm* (p)
E1039	*fms14* var. 1 (p)	*fms17* var. 1	*fms13* var. 1	*scm* var. 1 (p)	-	*fms15* (p)	*acm* (p)
1231408[f]	*fms14* var. 2[m]	*fms17* var. 2[n]	*fms13* var. 2[o]	*scm* var. 1 (p)	-	*fms15* (p)	*acm*
CA Clade							
1,141,733	*fms14* var. 2	*fms17* var. 2	*fms13* var. 2	*scm* var. 2 (p)[p]	-	-	*acm*
TX1330	*fms14* var. 2	*fms17* var. 2	*fms13* var. 2	*scm* var. 2	-	-	*acm* (p)
Com12	*fms14* var. 2	*fms17* var. 2	*fms13* var. 2	*scm* var. 2 (p)	-	-	*acm* (p)
Com15	*fms14* var. 2 (p)	*fms17* var. 2	*fms13* var. 2	*scm* var. 2 (p)	-	-	*acm* (p)
E980	*fms14* var. 2	*fms17* var. 2	*fms13* var. 2	-	-	-	*acm* (p)

[a] var. 1, variant 1; var. 2; variant 2.
Cut-off for var.2 was set at <95% aa identity versus TX16 using PBLAST.
[b](p), partial, incomplete gene sequence or pseudogene
[c] aa identities/similarities of proteins encoded by "*fms20*" var. 2 genes range between 39-43%/55-57% versus that of *fms20* var. 1 of TX16
[d] aa identities/similarities of proteins encoded by *fms16* var. 2 genes range between 93-94%/96-98% versus that of fms16 var. 1 of TX16
[e]*ebpA* and *ebpC* are fused into a single ORF in E1071 (5' region of *ebpA* (nt 1-2757) joined to 3' region of *ebpC* (nt 1083-1875))
[f] Strain 1,231,408 is a hybrid of HA and CA clades; see text for details
[g] aa identities/similarities of proteins encoded by *ebpA* var. 2 genes range between 85-90%/80-82% versus that of *ebpA* var. 1 of TX16
[h] aa identities/similarities of proteins encoded by *ebpB* var. 2 genes range between 87-92%/92-95% versus that of *ebpB* var. 1 of TX16
[i] aa identities/similarities of proteins encoded by *ebpC* var. 2 genes range between 90-94%/97-98% versus that of *ebpC* var. 1 of TX16
[j] aa identities/similarities of proteins encoded by *fms11* var. 2 genes range between 79-92%/85-95% versus that of *fms11* var. 1 of TX16
[k] aa identities/similarities of proteins encoded by *fms19* var. 2 genes range between 89-90%/94-95% versus that of *fms19* var. 1 of TX16
[l] aa identities/similarities of proteins encoded by *fms18* var. 2 genes range between 90-93%/94-96% versus that of *fms18* var. 1 of TX16
[m] aa identities/similarities of proteins encoded by *fms14* var. 2 genes range between 85-86%/92-94% versus that of *fms14* var. 1 of TX16
[n] aa identities/similarities of proteins encoded by *fms17* var. 2 genes are 91%/96% versus that of *fms17* var. 1 of TX16
[o] aa identities/similarities of proteins encoded by *fms13* var. 2 genes range between 66-68%/80-82% versus that of *fms13* var. 1 of TX16
[p] aa identities/similarities of proteins encoded by *scm* var. 2 genes range between 94-95%/96-97% versus that of *scm* var. 1 of TX16
[q] TC6 was left out of this analysis as it is a transconjugant, and therefore not a unique genome

Table from (Qin et al., 2012)

[3.3.7] Distribution of genes encoding MSCRAMM-like proteins, putative virulence genes, antibiotic resistance determinants, and CRISPRs.

Previous studies of *E. faecium* TX16 identified 15 genes encoding LPXTG family cell-wall anchored proteins with MSCRAMM-like features, such as immunoglobulin-like folding; 11 of these were found in four gene clusters, each predicted/demonstrated to encode a different pilus, and four were found as individual MSCRAMM-encoding genes (Nallapareddy et al., 2003; Sillanpaa et al., 2008; Sillanpaa et al., 2009a). Our search for these genes in 21 unique *E. faecium* draft genomes in this study found all of the MSCRAMM-encoding genes to be widely distributed except *fms18 (ecbA)* and *fms15* which were only in HA-clade isolates (although some are present as variants or pseudogenes within the HA-clade) (Table 17). Moreover, our analysis revealed that *ebpA-ebpB-ebpC*$_{fm}$, *fms14-fms17-fms13, fms20, scm*, and *fms18* (the latter present in only HA isolates) all have sequence variants in some of the 21 strains, with identities of the encoded variant proteins ranging from 39% (*fms20* homolog) to 94% (*ebpC*) versus their counterparts in TX16 (Table 17). In general, most of the MSCRAMMS followed the CA/HA clade groupings with a variant representing each clade. Variant 1 of the *fms11-fms19-fms16* locus was strictly found in the HA-clade, and variant 2 in the CA-clade except for 1,231,501 which only had one of the three proteins (*fms16*) as a CA-variant, suggesting recombination by this isolate. Variant 1 of *fms14-fms17-fms13* was found in all but one HA clade isolate (1,231,408, a hybrid of HA and CA clades, has variant 2) and variant 2 in all 5 CA-clade strains. Variant 1 of *scm* was found to be exclusively carried by all 16 HA clade strains and variant 2 by 4 of the 5 CA clade strains. Although the differences between these MSCRAMMs in CA- vs. HA-clade strains are generally greater (ranging from 2-27% with an average of 10%) than the differences (3-4%) previously reported for the clade-specific differences in a set of core genes that excluded predicted surface proteins, they are comparable to the differences seen in several other surface proteins that have been studied (Galloway-Pena et al., 2012; Galloway-Pena et al., 2011).

Interestingly, the majority of HA clade strains (12/16, including TX16) were found to have variant 1 of the *ebp* pilus operon, while variant 2 was exclusively found in the 5 CA-clade strains in addition to variant 1 in three of the five isolates. In contrast, variation within *fms20* was restricted to the HA clade; all CA clade isolates carried *fms20* variant 1, but the percent identity between these two variants is much smaller (39%), possibly indicating the need for a new gene name. Also of note was the *acm* gene, which is present as a pseudogene in all of the CA-clade isolates except 1,141,733 which is the only CA-clade isolate that is from a hospitalized patient; *acm* pseudogenes were also found in non-CC17 HA-clade isolates (Table 17).

We also searched the 21 *E. faecium* isolates for the presence and absence of 13 resistance genes. Our data correspond to previously published data for some of the isolates (Palmer and Gilmore, 2010; van Schaik et al., 2010) . We observed that there is a clear distinction between the isolates of the genetically defined CA clade and those of the HA clade with none of the CA clade isolates having any of the antibiotic resistance determinants analyzed (Table 18). On the other hand, all of the HA-clade isolates have multiple resistance determinants, including the *pbp5*-R allele that confers ampicillin

resistance previously reported by Galloway-Pena et al. (Galloway-Pena et al., 2011), except for strains 1,231,501 and E1039. 1,231,501, which is in the HA-clade but lacks all antibiotic resistances including *pbp5*-R, may have lost the allele via recombination and acquired *pbp5*-S or may even represent a more ancestral isolate. Indeed, 1,231,501 was shown to be a hybrid of HA and CA genomes by Palmer, et al., with the replacement (hybrid) region including *pbp5-S*, which could explain the origin of *pbp5-S* in this strain (Palmer et al., 2012). E1039, which has the *pbp5*-R allele but none of the other resistance genes, is genetically defined as a HA-clade isolate, but came from a healthy volunteer, perhaps explaining its lack of other antibiotic resistances. Interestingly, neither of these strains has IS*16*. D344SRF is the only other HA-clade isolate that lacks the *pbp5*-R allele; however, this strain is known to have spontaneously lost *pbp5* and the surrounding region and contains many other resistances (Rice et al., 2001). Of note, E1636 only has two of the 13 resistances analyzed (*tetM* and *pbp5*-R); however, this could possibly be explained by its early isolation in 1961. This again suggests that these isolates are more distantly related to the other strains within the HA-clade.

Two groups have previously analyzed CRISPR-associated genes within *E. faecalis* and *E. faecium* genomes (Palmer and Gilmore, 2010; van Schaik et al., 2010). Partial CRISPR-like loci were previously described in E1071, E1679, and U0317; however, these loci were within a gene and were considered non-functional (van Schaik et al., 2010). In addition, Palmer et al. identified CRISPR-cas predicted proteins in the Broad Institute strains Com12; 1,141,733; and 1,231,408 (Palmer and Gilmore, 2010). Similarly, we only found a CRISPR-cas locus in strain TX1330 (Table 19) out of the 6 strains not previously studied (TX1330; TX16; TX82; TX0133A; D344SRF; and C68). In summary, out of the 22 available genomes, only one of the HA-clade isolates contained CRISP-loci, namely the hybrid strain 1,231,408. The three other strains containing CRISPR-loci of the CA-clade (Com12; 1,141,733; and TX1330) all lacked antibiotic resistance determinants. Therefore, our data coincide with the previous observation that members of the recently emerged high-risk enterococcal lineages lack CRISPR-loci and the inverse relationship between the presence of a CRISPR-cas locus and acquired antibiotic resistance (Palmer and Gilmore, 2010). (Qin et al., 2012)

TABLE 18. Antibiotic resistance gene profiles of the 21 *E. faecium* strains

Gene	*cat*	*ermA*	*ermB*	*aad6*	*aad9*	*aadE*	*aacA-aphD*	*tetL*	*tetM*	*vanA*	*gyrA*[b]	*parC*[c]	*pbp5*-R[d]
Resistance	CHL	ERY	ERY	SPC/ STR	SPC/ STR	SPC/ STR	GEN	TET	TET	VAN	CIP	CIP	AMP
Strains													
1,141,733													
Com12													
Com15													
E980													
TX1330													
1,230,933			X	X		X	X		X	X	X	X	X
1,231,408			X	X		X	X				X	X	X
1,231,410			X	X		X				X	X		X
1,231,501													
1,231,502			X	X		X	X			X	X	X	X
C68			X	X		X	X		X		X	X	X
D344SRF[a]			X	X		X		X	X				
TX16	X		X	X		X		X	X				X
E1039													X
E1071	X		X	X	X	X		X		X			X
E1162								X	X				X
E1636									X				X
E1679		X	X	X	X		X			X	X	X	X
TX82			X	X		X			X	X	X	X	X
TX0133A	X		X	X		X	X		X		X	X	X
U0317			X	X		X	X				X	X	X

[a] A rifampin- and fusidic acid-resistant derivative of clinical strain *E. faecium* D344S in which the spontaneous loss of *pbp5* and its surrounding region resulted in an ampicillin-susceptible phenotype

[b] Amino acid change (E to K/G) in residue 87 or (S to R/Y/I) in residue 83 of GyrA

[c] Amino acid change (E to K) in residue 86 or (S to R/I) in residue 82 of ParC

[d] Consensus sequence of the *pbp5*-R allele encoding the low affinity Pbp5-R

[e] TC6 was not included in this analysis as it is a transconjugant of C68 and D344SRF, so therefore is not a unique genome

Table from (Qin et al., 2012)

TABLE 19. Summary of CRISPRs found in *E. faecium* sequenced strains.

E. faecium strain	Locus Tag	Conserved domain/ functional assignment
TX1330	HMPREF0352_0914	TIGR01865/*csn1*
	HMPREF0352_0915	TIGR03639/*cas1*_NMENI
	HMPREF0352_0916	COG3512/*cas2*
	HMPREF0352_0917	pfam09711/*csn2*
Com12[a]	EFVG_01552	TIGR01865/*csn1*
	EFVG_01553	TIGR03639/*cas1*_NMENI
	EFVG_01554	COG3512/*cas2*
	EFVG_01555	pfam09711/*csn2*
1,141,733[a]	EFSG_02406	TIGR01865/*csn1*
	EFSG_02407	TIGR03639/*cas1*_NMENI
	EFSG_02408	COG3512/*cas2*
	EFSG_02409	pfam09711/*csn2*
1,231,408[a]	EFUG_01572	TIGR01865/*csn1*
	EFUG_01573	TIGR03639/*cas1*_NMENI
	EFUG_01574	COG3512/*cas2*
	EFUG_01575	pfam09711/*csn2*

[a] Previously published (Palmer et al., 2010)

Table from (Qin et al., 2012)

[3.4] CHAPTER 2 DISCUSSION

Many publications have emphasized the importance of the accessory genome in distinguishing the two subpopulations of *E. faecium*, and have often assumed that the accessory genome was responsible for the ability to adapt to the nosocomial environment (Leavis et al., 2007; van Schaik et al., 2010; van Schaik and Willems, 2010). However, in the previous chapter a considerable intra-species difference (3-10%) was noted in the sequences of *pbp5*, *pbp2*, *wlcA*, and *gls20*, with each of the gene sequences forming two distinct groups generally corresponding to HA and CA strains (Galloway-Pena et al., 2011). This led to our hypothesis that differences between the HA and CA groups may be at a more fundamental level (i.e., in the core genome). An extensive analysis determining the degree of the core genomic differences using the numerous *E. faecium* draft genome sequences and one closed *E. faecium* genome sequence had not been reported.

Our analysis presented above showed a difference in the concatenated sequences of 100 genes between the two clades of approximately 3.5-4.2%. Greater than 90% of the 100 core genes analyzed separated into two distinct groups, representing a variety of metabolic and cellular processes (Table 10 and Figure 7). Interestingly, ~60% of the sequence changes seem to be clade-specific changes (sequence changes found in the majority of the strains in one clade, but not the other), while the rest were strain specific. Even though the 21 whole genome analysis confirmed that there are differences in the accessory genomes of HA and CA strains (e.g. putative virulence genes, antibiotic resistance determinants, IS elements and transposons), the changes in genes encoding core metabolism and cellular processes could be another reason why some strains adapt better to the clinical environment.

Our comparison of the 21 *E. faecium* genomes found that the majority of the 15 MSCRAMM and pilus encoding genes described in Chapter 1 (12/15) were enriched in HA

clade strains. If they were present in all strains, they had a different sequence variant associated with each clade, suggesting that these genes may have also contributed to the emergence of HA-clade *E. faecium* in the clinical setting. Also of note, all CA clade strains lacked all antibiotic resistance determinants analyzed in this study. Three of these CA clade strains contained the CRISPR-cas locus, supporting the hypothesis that there is an inverse correlation between the presence of CRISPR loci and acquired antibiotic resistance (Palmer and Gilmore, 2010). Although IS*16* presence was highly correlated with HA strains, it was not found in all HA-clade strains. Interestingly, however, all HA-clade strains did contain the *pbp5*-R allele discussed in Chapter 1. The *pbp5*-R allele is also found in some animal and community isolates that are genetically within the HA-clade, but are not clinically associated, possibly arguing that it is a more reliable marker for HA-clade associated strains. The exceptions are D344SRF, which is a spontaneous deletion mutant of *pbp5*, and 1,231,501. 1,231,501 is a HA-clade isolate from the blood of a hospitalized patient, but has no resistance genes (contains the *pbp5*-S allele), supporting the hypothesis that the genomic content of a strain, not just antibiotic resistance, adds to the survival in the hospital environment.

However, not all strains that grouped genetically within the HA or CA clades had the indicated origin. For example, the hospitalized patient blood culture isolate, 1,141,733, always grouped with strains in the CA clade, whereas E1039, a healthy volunteer fecal sample, belongs to the HA clade. 1,231,408 did not fall strictly into one clade, so it was categorized as a hybrid strain. The SNP analysis determined that this strain recombined around ORF 10683, as its SNP sequence prior to ORF 10683 showed near identity with the concatenated SNP sequence of CA clade strains, whereas, after ORF 10683, its concatenated sequence showed close identity to the SNP sequence of HA strains.

The general notion held by most was that HA isolates had evolved from CA isolates under the pressure of antibiotic use and the environment of the hospital setting. We tested this idea by estimating the time of separation of the HA and CA clades since the hypothesis of cladal separation held true for the majority of the isolates. Two main strategies used in a number of other publications, 16S rRNA and synonymous SNPs throughout the genome, were used to analyze the molecular evolution of *E. faecium* in this study (Achtman et al., 2004; Foster et al., 2009; Lenski et al., 2003; Ochman et al., 1999; Ochman and Wilson, 1987; Van Ert et al., 2007). Although 16S rRNA analysis is historically the accepted methodology to determine speciation, recent studies have expressed concerns regarding 16S rRNA being a reliable chronometer. sSNP sites of protein-encoding genes are not affected by selection or genetic drift and therefore are thought to reflect the underlying rate of mutation more reliably than 16S rRNA (Kuo and Ochman, 2009; Pearson et al., 2009) and to be more useful for species with high levels of recombination, such as *E. faecium* (Pearson et al., 2009). Therefore, both 16S rRNA and sSNP analysis were used. The 16S rRNA analysis shows a 0.06% to 0.1% difference between the two clades (Figure 11 and Table 15) which puts the time of divergence between 1.5-2.5 million years ago. The molecular clock estimate using sSNPs analysis and the mutation rate of *E. coli* has the strains within the HA clade diverging ~100,000 to 300,000 years ago. Using the same formula, strains within the CA clade probably diverged from each other ~300,000 to 900,000 years ago. Even sSNP analysis using the mutation rates up to 1000 fold higher estimated for *B. anthracis* estimated divergence time between the two clades at approximately 300,000 years ago.

These data together (core genes, SNPs and 16S rRNA) indicate a large divergence between CA-clade isolates and HA-clade isolates. Although the estimates for the time of

divergence are crude, they indicate that the CA clade and HA clade isolates diverged long before the modern antibiotic era. Nonetheless, the molecular clock calculations do support the hypothesis that HA clade strains stem from a relatively recent ancestor and are the more recently evolved genogroup. This analysis highlights that core genomic differences, in addition to the accessory genomic differences, may be why some strains adapt to the nosocomial environment and become opportunistic pathogens, while others do not.

Using these data as well as other previous publications (Galloway-Pena et al., 2012; Galloway-Pena et al., 2011; Leavis et al., 2007; Palmer et al., 2012), scenarios for the evolution of the two clades of *E. faecium* are depicted in Figure 14. One may hypothesize that a primordial type of *E. faecium* split many millennia ago into two early groups with distinct differences between their genes (i.e. the *pbp5*-S or *pbp5*-R alleles). These lineages could recombine with each other resulting in hybrid strains (i.e. 1,231,408 and 1,231,501) (scenario 1). These lineages continued to evolve, eventually reaching a core genomic difference of approximately 3-4%, creating a HA clade, which includes AREfm community-based isolates as well as most of the hospital-associated isolates (scenario 2). This clade is separate from the CA clade consisting of mostly community-derived isolates. A few scenarios exist for the evolution of AREfm clones. One scenario, is that community and hospital AREfm isolates split from the same ancestor (scenario 2). However, it is possible that AREfm clones evolved from the animal reservoir (scenario 3) or that the AREfm isolates found in animals are HA clade AREfm transferred from humans to their pets (scenario 4).

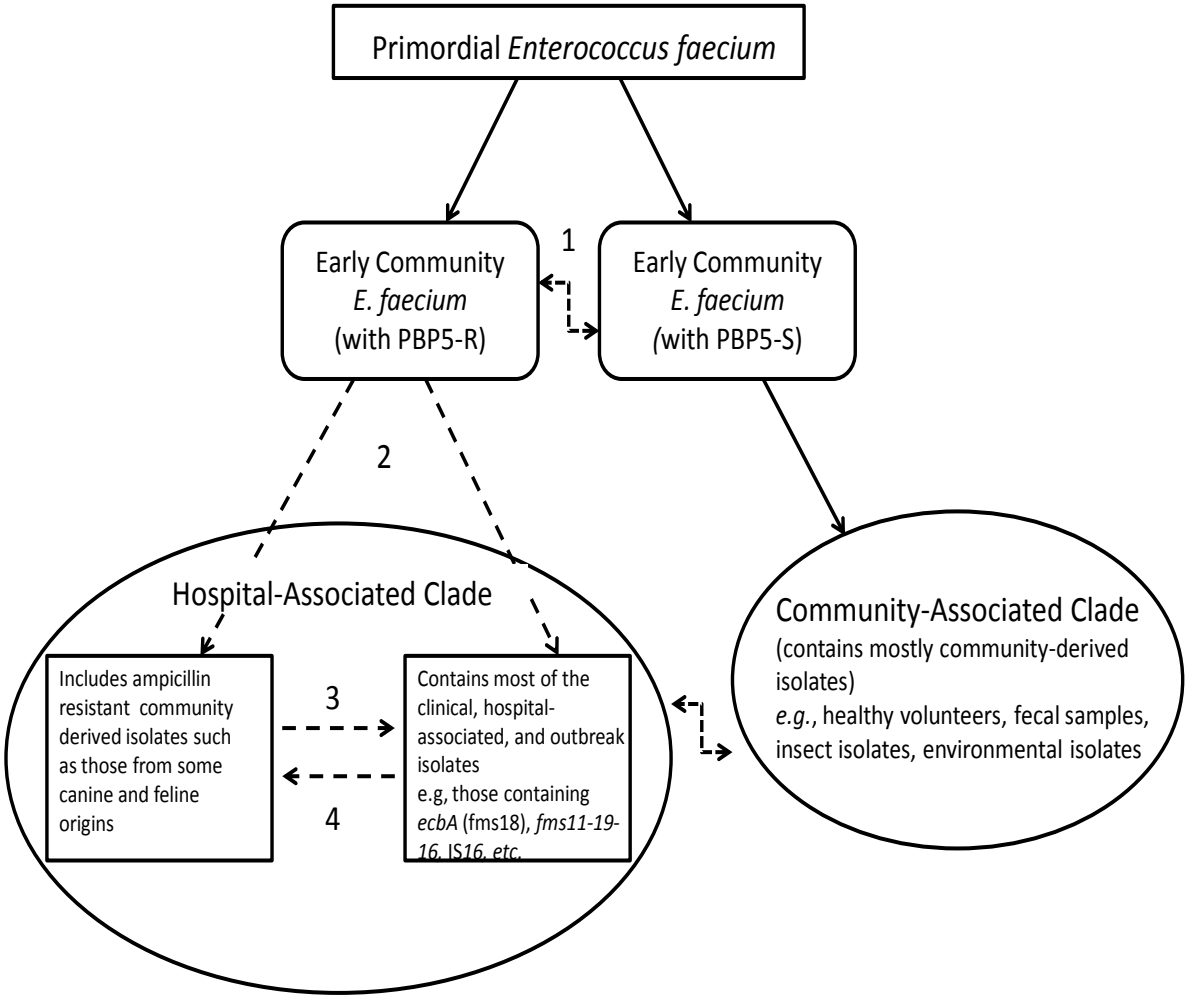


FIGURE 14. The projected evolution of the two clades of *E. faecium*. A figure addressing the possible scenarios that may have occurred in the evolution of *Enterococcus faecium* resulting in the HA-clade and CA-clade. Specifically, a primordial type of *Enterococcus faecium* split into early community isolates which had homologous core genomes with significant sequence differences (e.g., the *pbp5*-S or *pbp5*-R allele). These early community groups further segmented into a hospital-associated clade and the community clade. Scenario one depicts that these lineages could recombine with each other (represented by the bent dashed

arrow) resulting in hybrid strains, scenario two depicts community and hospital AREfm isolates splitting from the same ancestor, scenario three depicts AREfm clones evolving from the animal reservoir, and scenario four depicts animal AREfm isolates representing descendants of hospital ARE transferred from humans to their pets. Figure from (Galloway-Pena et al., 2012).

[4] CHAPTER 3

Defining the Role of "WxL" Proteins in *Enterococcus faecium*

[4.1] CHAPTER 3 INTRODUCTION

With the availability of numerous draft genomes and one closed genome (TX16) of *E. faecium*, we performed a search throughout these genomes looking for MSCRAMMS (discussed in the Background section and in Chapter 2) and other surface proteins that might be potential virulence factors contributing to the initiation of disease. One newly described class of surface protein was included in this search, called the "WxL" proteins.

Genes encoding proteins with the "WxL" motif were first described in *Lactobacillus plantarum.* A novel gene cluster thought to encode cell-surface proteins was identified and found to be conserved in a subset of Gram positive bacteria. Proteins encoded within each cluster had a signal peptide for secretion via the Sec-dependent pathway. Some of these had cell-surface anchoring motifs similar to MSCRAMMs, and others contained a novel "WxL" domain (each cluster contained 2 of these). In addition, putative domains for sugar binding and degradation, CRE-sites upstream or inside the gene clusters, and transcriptome analysis showing higher expression levels of these gene clusters in a glucose-grown *ccpA* (catabolite control protein A) mutant, suggested that these proteins formed cell surface protein complexes involved in carbon source acquisition and stress survival adaptation. However, all of these data were from *in silico* genomic and transcriptome analysis, and none of the speculations were experimentally proven (Siezen et al., 2006).

In another publication, a homolog of the proteins containing the WxL domain found in *Lactobacillus coryniformis* was purified, and shown to mediate co-aggregation with itself and

other pathogens like *E. coli* and *C. jejuni*. It was therefore named Cpf (Co-aggregation promoting factor). Cpf could be removed from the surface of the cell and released into the supernatant using 5M LiCl, suggesting a non-covalent attachment, and then reattached to the cell surface to restore the co-aggregation phenotype. It was suggested that the co-aggregation phenotype found in this publication had a competition effect in the gastrointestinal tract by preventing other pathogens from adhering to host cells (Schachtsiek et al., 2004).

In 2007, the clinical strain *E. faecalis* V583 was analyzed for novel cell surface proteins. Twenty-seven of the proteins found possessed the C-terminal WxL domain earlier described in lactobacilli. By generating fusion constructs between the C-terminal region of two of the WxL proteins (using only the WxL domain) and a nuclease reporter gene, the authors were able to show that the WxL domain conferred localization to the cell wall fraction. The localization was eliminated or lessened when this domain was absent or truncated. These recombinant protein fusions were also shown to bind other Gram positive bacteria, supporting the coaggregation findings with Cpf discussed above. Furthermore, it was shown that peptidoglycan was the binding ligand for the WxL domain attachment to the surface. Therefore, the WxL domain itself was considered a novel cell-wall binding domain in *E. faecalis* (Brinster et al., 2007a).

The same group took the *E. faecalis* studies further by deleting one of the 27 genes encoding WxL proteins, named ElrA (enterococcal leucine-rich protein A). This gene was chosen because it had both a WxL domain and a leucine-rich repeat (LRR) domain which is characteristic of proteins (internalins) from *Salmonella*, *Shigella*, and *Listeria* spp. involved in host-pathogen interactions and host cell invasion. Disruption of the *elrA* gene lead to significant attenuation in a mouse peritonitis model and reduced bacterial dissemination to the spleen,

liver, and host macrophages. Furthermore, RT-PCR analysis showed that *elrA* gene expression was induced in vivo (Brinster et al., 2007b).

The research contained within this chapter aimed to tests the hypothesis that proteins containing the WxL domain are important for virulence/colonization in *E. faecium* using the following specific aims: 1) To determine the localization of WxL proteins of *E. faecium* and the possible interactions of the proteins encoded by the same locus, 2) To determine if any of the WxL operons affect colonization or virulence associated functions, and 3) To determine if WxL proteins are antigenic during natural infection and if antibodies to the WxL proteins confer a protective role.

[4.2] CHAPTER 3 MATERIALS AND METHODS

TABLE 20. Bacterial strains and plasmid used in this study.

Strain or Plasmid	Relevant Characteristics	Reference
TX1330	*E. faecium* fecal isolate from a healthy volunteer	(Qin et al., 2012)
TX1330RF	Rifampin and fusidic acid resistant version of TX1330	(Arias et al., 2009)
TX82	Endocarditis isolate Van^R, Erm^R	(Nallapareddy et al., 2006)
M15(pREP4)	*E. coli* strain for expression of recombinant proteins, Kan^R	Qiagen
pQE30	Recombinant protein expression vector Amp^R	
TX6116	M15 containing fragment of $DufA_{45-309}$ from TX82 cloned into pQE30	Qiagen
TX6117	M15 containing fragment of $LwpC_{10-562}$ from TX82 cloned into pQE30	This Study
TX6118	M15 containing fragment of $LwpA_{33-716}$ from TX82 cloned into pQE30	This Study
TX6119	M15 containing fragment of $SwpA_{26-226}$ from TX82 cloned into pQE30	This Study
TX6120	M15 containing fragment of $SwpC_{25-243}$ from TX82 cloned into pQE30	This Study
TX6121	M15 containing fragment of $LwpC_{35-587}$ from TX1330 cloned into pQE30	This Study
TX6122	M15 containing fragment of $SwpC_{26-243}$ from TX1330 cloned into pQE30	This Study
TX6123	M15 containing fragment of $LwpA_{33-772}$ from TX1330 cloned into pQE30	This Study
TX6124	M15 containing fragment of $SwpA_{26-227}$ from TX1330 cloned into pQE30	This Study
CK111	*E. faecalis* strain used as a donor for mating in mutant production	(Kristich et al., 2007)
pHOU1	pCJK47 derivative used to make markerless non-polar deletions in *E. faecium*, Gen^R	(Panesso et al., 2011)
TX5725	CK111 containing WxL Locus B constuct to create a deletion in TX82	This Study
TX6125	CK111 containing WxL Locus A constuct to create a deletion in TX82	This Study
TX6126	CK111 containing WxL Locus C constuct to create a deletion in TX82	This Study
TX6106	WxL Locus A deletion in TX82	This Study
TX6107	WxL Locus C deletion in TX82	This Study
TX6115	WxL Locus B deletion in TX82	This Study

[4.2.1] Strains, plasmids, and cultivation of bacteria. All relevant characteristics of bacterial strains, antibiotic growth conditions, and plasmids used in this study can be found in Table 20. Unless otherwise stated, all *E. coli* cells were grown in LB broth/agar and all *E. faecalis* or *E. faecium* cells in BHI agar and broth at 37C.

[4.2.2] Sequencing of loci and *in silico* genomic analysis. The genome sequence of TX16 was first searched for ORFs with the potential to encode proteins with the WxL domain using NCBI BLAST. Initially, ORFs were predicted with Glimmer 2 and protein domains determined with InterProScan. All ORFs generated through BLAST longer than 150 amino acids were then searched for the "WxL" motif (and other conserved residues; see section 4.3.1. Figure 15) within the C-terminal 150 amino acid residues. The WxL domain was confirmed using CLUSTALW alignment with previously predicted WxL proteins (Siezen et al., 2006). The gene organization of the ORFs surrounding the predicted "WxL" ORFs were also analyzed to determine if they resembled what was previously reported (Siezen et al., 2006). InterProScan was used to determine protein family, domains, and functional sites. TMHMM v2.0 and TMpredServer were used to determine possible transmembrane domains. Signal P v3.0 was used to determine the presence of N-terminal signal sequences. SoftBerryBPROM and SoftBerryFindTerm as well as RNA secondary structure predictor (www.genebee.msu.eu) were used to determine the predicted promoters and terminators of the loci. Secondary and tertiary structures were predicted using PHYRE. DisEMBL v1.5 was used to predict intrinsic protein disorder. NCBI BLAST was then used to find the WxL loci in TX1330 and TX82 using the sequence from TX16. Confirmatory sequencing of the WxL loci using the genomic DNA from the *E. faecium* strains TX1330 and TX82 were performed using the primers found in Table 21.

TABLE 21. Primers used for the re-sequencing of the WxL loci.

Oligonucleotide[a]	Sequence 5′-3′
Locus A	
WxL27seqF1	CACAGCATCCATCGTTTGGC
WxL27seqR1BBB	CAGCCACGAGGTTCCAAGTCAAATC
WxL27toDUF916Fwd	GTAAAGCAAGTGCAGCGACACC
WxL27toDUF916Rev	CGCCGGATTAACAGAGACCTCAAATG
DUF916seqF1	GTAACGAGTGATGGTGGAACGATC
DUF916seqR1	CAACGAAATCACCTCATCG
DUF916toWxL16Fwd	CAGGTGACAGTAATGAATGGCG
DUF916toWxL16Rev	CAGTTAGTTGGACAAGCTTCCC
WxLHyb1and6Fwd	TTCGGCAGAAGCTGTTGGAGCAA
WxLHyb1and6Rev	CCACGTATTTATCTCTGCACTTGCAG
WxL16seqF2	GCTCCAGCTGCAGGTAGATTTTATC
WxLseq16R2BBB	CTGAAGCCAATGCAGGTAGTGTAG
WxL16seqF3	CTACACTACCTGCATTGGCTTCAG
WxL16seqR3BBB	GTTGCCAAGGAGACGTGCTTGGAAG
WxL16seqF4	CAATTCTTGTCCCCAAACAAG
WxL16seqR4	GGTAGACCGATCGATTGATAAGCG
Hypo1seqF1	GGCCAATCAGCTAAGTGATCAG
Hypo1seqR1	CGATTCAATTGAGCCACTACGTCATC
Hypo1seqF2	CCATTATGAAGAAACGGATGCGCTG
Hypo1toOx2Rev	GTCACAGAAGTCCGCTAAAG
Hypo1toOx2Fwd	CCTCAGAAGTGAGGATGAGCTGATG
Ox2seqR1	GACCATCACAGTGTCGTCCTCATAC
Ox2seqF1	GGGTTAGAATTAGCAGATGC
Ox2seqR2	GCTTATCAGCTGTCCACCAATG
Ox2toABCtransFwd	CTGTTGGGTTGACAGAAGACGAAG
Ox2toABCtransRev	GACTCCTGGAGTTTGGGCAAGATAG
Locus B	
WxL410seqF1	AAACGTCAGAGAGGATGGTG
WxL410seqR1	CTGTTTCCCCAAAAATCGGA
WxL 410 F #2	CGATGGAAATGCCATTATCATTGC
WxL 410 R # 2	CTCCAGCCATTTTGACCGAGATTTGTCC
WxL410seqF2	GCCTGCAAGGAGATTCTAG
WxL410seqR2	TTCATTCCGGCACCTGTCTGAGC
WxL410seqF3	CGATCAATGGAACCGACTTAG
WxL410seqR3	TCTTGCGTCAACCCAGTTGC
WxL410toCWAFwd	CCAAACTCGATTGTGGTCAGATC
WxL410toCWARev	GAGGAGGAACAGTTATTTCTCCCAC
CWAseqF1	CGGTGTGATCGAATATGGAG
CWAseqR1	GGCTGAGCTCTTATTCTCAG
WxLHyb3and11Fwd	GCATTAAGTACCGTTCTAGCTGG
WxLHyb3and11Rev	CTGTAGGAGCATCTGTTAACGACC
WxL311seqF1	GCCGTAGACAACTACGTTTC

WxL311seqR1	GTGGAAGTCTCGTCCTGTATCCTC
Up410F1	CATTCTCCACCATCCTCTCTGACG
Up410R1	GTTCGGGATAAGGGACTGTGACAAGAC
Locus C	
WxL5A9seqF1	GAAGGAAACCGCGACACCTATTG
WxL5A9seqR1	GATCTTCCGGATCAACTGGAGGAG
WxL 5A9 F #2	ACATTTGAAGCAGGAGATGAGGG
WxL 5A9 R #2	AAGTCCCCACACCTGTTCCTGATTTAGC
WxL5A9toWxL5B8Fwd	GCTAAATCAGGAACAGGTGTGG
WxL5A9toWxL5B8Rev	CCAAATAGCTCCTCTTTGAGACC
WxLHyb5Band8Fwd	GGATCGTGCTCATTTTAAGCAG
WxLHyb5Band8Rev	CCATTCGATCAGAAGAACCACTA
WxL5B8seqF1	CGCAAATGCACTCTCGATTG
WxL5B8seqR1	ACACCTTCTGTATTCGCTGA
5B8toDUF916#2Fwd	GATTCAGCGAATACAGAAGG
5B8toDUF916#2Rev	CGTTTGAATTAGTCGTTGCGC
DUF916seqF1	CGATCTCCAACTTGTTCCAG
DUF916R1	CAGCGCTTCTAAGATGTAGATCC
DUF916toCWA0757Fwd	GGATCTACATCTTAGAAGCGC
DUF916toCWA0757Rev	GCGAGATAATAGGGCATTACTAG
CWA0757seqF1	CTGTTAGTGAATTGGTGATATGTTTGG
CWA0757seqR1	TATCTAGTGGAAATTCATCC
CWA0757seqF2	CGGTTACCGAAAGAAACGAG
CWA0757seqR2	CACCACGTCTTGTACACGCACTACTTC
WxL5A9seqR1BBB	CCCTCATCTCCTGCTTCAAATG
Oligos for Gaps in Sequencing	
LA1401-1902F	GAAGAAATAAGAGAAGCAGGCGGG
LA1401-1902R	CGCTGTCAAACACGATTGGTAC
LA1857-2463F	GTGGCAGAAGCAGTAGATAAACG
LA1857-2463R	GCCGCCTAGAAGAAGTGTTGAAAG
LA2440-2957F	CTTTCAACACTTCTTCTAGGCGGC
LA2440-2957R	GCTGCTGTAAATATTGGTGTCGCTGC
LA3352-3755F	CCGGAACAAAATCAGACACTGG
LA3352-3755R	GATCCGGTTCTATCGTCTCTTGAC
LA5964-6807F	CGATCTCATCAAATAATGGCGGC
LA5964-6807R	CCGCTTAATTCTTGACCTTGATC
LA7201-7824F	GAGGAAAAAGGAGAGATCTCTG
LA7201-7824R	CAATCCAACAGGAAGATTAC
LA9690-10370F	GTGCCTGTTCTTGTTCAGGAGAG
LA9690-10370R	AGAACGATCGAATCTGCCTC
LA10321-11024F	CAAGACAATCGAACGGCGGTTGAC
LA10321-11024R	CCTTGGGCTTTCATTACACC
LB5081-5665F	CTTTCCCTTCTGTAGGAGCATCTG
LB5081-5665R	GCAGTAGAGATCATCAATCGG
LB5645-6120F	CCGATTGATGATCTCTACTGC

LB5645-6120R	GTCACCACCATCAAAGCTATACAACC
LC496-1020F	GCTGTCTTATGATGCAACCATTCC
LC496-1020R	CCATCGCTCCCGTAGTAATAAC

[a]WxL proteins were named differently when experiments first began. WxL27 is SwpA, WxL16 is LwpA, WxL410 is LwpB, WxL311 is SwpB, WxL5A9 is SwpC, WxL5B8 is LwpC.

[b] LA is Locus A, LB is Locus B, and LC is Locus C

TABLE 22. Primers for probes used in hybridizations of genes encoding WxL proteins.

Oligonucleotide[a]	5'-3' Sequence	Size of Probe
WxLHyb16Fwd	TTCGGCAGAAGCTGTTGGAGCAA	
WxLHyb16Rev	CCACGTATTTATCTCTGCACTTGCAG	798bp
WxLHyb27Fwd	TTCTTCTAGGCGGCTTGGCAA	
WxLHyb27Rev	AGCCACGAGGTTCCAAGTCAAA	625bp
WxLHyb311Fwd	GCATTAAGTACCGTTCTAGCTGG	
WxLHyb311Rev	CTGTAGGAGCATCTGTTAACGACC	729bp
WxL 410 F #2	CGATGGAAATGCCATTATCATTGC	
WxL 410 R # 2	CTCCAGCCATTTTGACCGAGATTTGTCC	679bp
WxL 5A9 F #2	ACATTTGAAGCAGGAGATGAGGG	
WxL 5A9 R #2	AAGTCCCCACACCTGTTCCTGATTTAGC	474bp
WxLHyb5B8Fwd	GGATCGTGCTCATTTTAAGCAG	
WxLHyb5B8Rev	CCATTCGATCAGAAGAACCACTA	669bp

[a]WxL proteins were named differently when experiments first began. WxL27 is SwpA, WxL16 is LwpA, WxL410 is LwpB, WxL311 is SwpB, WxL5A9 is SwpC, WxL5B8 is LwpC.

[4.2.3] Colony hybridization. Preparation of colony lysates on nylon membranes and hybridization under high-stringency conditions were performed as described previously (Coque et al., 1995). Briefly, *E. faecium* isolates were replica plated on nylon membranes from 96-well microtiter plates and then placed on BHI plates and grown overnight at 37C. The colonies were lysed using a treatment with lysozyme (10mg/ml) and mutanolysin (4units/ml) for 30 minutes at 37C. The genomic DNA left on the membrane was denatured with 0.5M NaOH for 10 minutes at room temperature. The membranes were then washed with 1M Tris (pH7.6), 0.5M Tris (pH7.6) + 1.5M NaCl, chloroform, then 95% EtOH. The DNA was then fixed by baking

the membranes at 80C for 4 hours. Hybridizations were performed under high stringency conditions using radiolabelled probes (RadPrime DNA Labeling System Invitrogen) in hybridization buffer containing 50% formamide, 5× Denhardt's solution, 5× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate, pH 7.2), 0.1% SDS, and 100 μg/ml calf thymus DNA at 42°C overnight, followed by three washes with 2× SSC and 0.1% SDS at room temperature (15 min each) and two washes with 0.1× SSC and 0.1% SDS at 50°C (15 min each). DNA probes for the WxL loci genes were obtained with primers listed in Table 22.

[4.2.4] RNA extraction and RT-PCR. Total RNA was isolated from TX1330 and TX82 grown to mid-log phase in 10mL of BHI using RNAwiz according to manufacturer's instructions (Ambion). Briefly, 500μl of RNAwiz was added to the bacterial pellet, and the mixture vortexed using 0.5g of silica beads for 5 minutes. After centrifugation, an additional 500μl of RNAwiz was added and vortexed again. 300μl of chloroform was then added and incubated at RT for 10 minutes. The mixture was then centrifuged at 4C for 15 minutes, and the supernatant collected. The RNA was then ethanol precipitated. The RNA pellets were resuspended in 100μL of water and treated twice with 20U RQ1 DNase (Promega) for 1 hour at 37C. DNase was then removed using the RNeasy Mini kit purification protocol (Qiagen). The total RNA was then reverse-transcribed with the specific primers found in Table 23 using the SuperScript One-Step RT-PCR with Platinum Taq kit (Invitrogen) using the manufacturer's instructions. Reactions without reverse transcriptase were used to verify that there was no DNA contamination and reactions with genomic DNA were used to verify the effectiveness of the primers.

TABLE 23. Primers used for RT-PCR of the intergenic regions of WxL loci.

Oligonucleotide[a,b]	5'-3' Sequence
Locus A TX1330	
Ox1toWxL27Fwd	GGTCATGCAATTAGCTGGAACG
Ox1toWxL27Rev	CCAAACGGCAATTCGGGAACTAC
WxL27toDUF916Fwd	GTAAAGCAAGTGCAGCGACACC
WxL27toDUF916Rev	CGCCGGATTAACAGAGACCTCAAATG
DUF916toWxL16Fwd	CAGGTGACAGTAATGAATGGCG
DUF916toWxL16Rev	CAGTTAGTTGGACAAGCTTCCC
WxL16toHypo1Fwd	CCAACAAGGACAAAACACGATG
WxL16toHypo1Rev	CCATAACAAGTTTCTGCCGATC
Hypo1toOx2Fwd	GTCACAGAAGTCCGCTAAAG
Hypo1toOx2Rev	CCTCAGAAGTGAGGATGAGCTGATG
Ox2toABCtransFwd	CTGTTGGGTTGACAGAAGACGAAG
Ox2toABCtransRev	GACTCCTGGAGTTTGGGCAAGATAG
Locus C TX1330	
Aldolase to WxL5A9F	GATGAATACACAAGGACGAG
Aldolase to WxL5A9R	CGAAAGTCGTACTTACTGCC
WxL5A9toWxL5B8Fwd	GCTAAATCAGGAACAGGTGTGG
WxL5A9toWxL5B8Rev	CCAAATAGCTCCTCTTTGAGACC
5B8toDUF916#2Fwd	GATTCAGCGAATACAGAAGG
5B8toDUF916#2Rev	CGTTTGAATTAGTCGTTGCGC
DUF916toCWA0757Fwd	GGATCTACATCTTAGAAGCGC
DUF916toCWA0757Rev	GCGAGATAATAGGGCATTACTAG
CWA to Glycerol MIP F	CCAAGAAGAGAAATACAAAGACAGC
CWA to Glycerol MIP R	GGTACCCGTTGTTGCACCTATCTGTGC
Locus A TX82	
TX82Hydro to WxL27 F	GTTTGACAGCGGTGTACGAC
TX82Hydro to WxL27R	GCTGGCTCTACTGGTTTTGAAG
WxL27toDUF916Fwd*[c]	GTAAAGCAAGTGCAGCGACACC
TX82WxL27 to LA DUF R	CCACCATCACTCGTTACCGC
DUF916toWxL16Fwd*	CAGGTGACAGTAATGAATGGCG
DUF916toWxL16Rev*	CAGTTAGTTGGACAAGCTTCCC
TX82WxL16 to TransRegF	GATCAAGGTCAAGAATTAAG
TX82WxL16 to TransRegR	TGCTGCTATATTTCCATGTCTG
Hypo1toOx2Fwd*	GTCACAGAAGTCCGCTAAAG
TX82TransReg to Ox R	GAGGATGAGCTGATGTTTGTCTGG
TX82Ox to ABC F	CAAAAC AGTAAATGCA AAGC
TX82Ox to ABC R	CTTCTTCTGTACTGTCTTTACTAG
Locus B TX82	
TX82Kinase to WxL410 F	GCAAAAGTGGATGATGGTAAG
TX82Kinase to WxL410 R	CTTGTCTTATTTTCTTCTTGCC

TX82WxL410 to DUF916 F	CAAGAAAGGAGAGGAGTATCAA
WxL410toCWARev*	GAGGAGGAACAGTTATTTCTCCCAC
CWAtoWxL311Fwd*	GGGACATTATCAAGCGGTTGTG
TX82DUFtoWxL311 R	GTAGTCAAGCCGAACATATAAGAAGC
TX82WxL311toTransReg F	CAAAGGTGAAATTCAAAAACGCC
TX82WxL311toTransReg R	GATGAATAAGAGCTGAGACGTAAG
Locus C TX82	
TX82Metalo to WxL5A9 F	GTTTTTGGCCATACGCATCGTCGC
TX82Metalo to WxL5A9R	CCATAATCAATCGATAATGCCCCTCC
WxL5A9toWxL5B8Fwd*	GCTAAATCAGGAACAGGTGTGG
WxL5A9toWxL5B8Rev*	CCAAATAGCTCCTCTTTGAGACC
WxL5B8toDUF916Fwd*	GGATATCTATTTTTATTACAG
TX82WxL5B8toDUF916 R	CTTCTTTTTCTGTATCATTTC
TX82DUF to CWA F	AGGAAATGAAAATGAAAATGGA
TX82DUF to CWA R	CCTGAAAGAAAGCCAATACC
TX82CWA to MIP F	GAATAGGAATAATCTATATGAAGAAAG
TX82CWA to MIP R	TTCTTTGGTTCATTCGTTTTATTC

[a]WxL proteins were named differently when experiments first began. WxL27 is SwpA, WxL16 is LwpA, WxL410 is LwpB, WxL311 is SwpB, WxL5A9 is SwpC, WxL5B8 is LwpC.

[b] Oligonucleotide pairs are listed in the order in which they were used for each Locus to amplify intergenic regions, i.e. the gene order

[c] Asterisks indicate an oligonucleotide used for the RT-PCR of both TX1330 and TX82

[4.2.5] Construction of expression plasmids and purification of recombinant proteins. The genomic DNA of *E. faecium* was extracted from TX1330 and TX82 cells grown overnight at 37C in 10 mL of BHI using DNAeasy Blood and Tissue Kit (Qiagen). DNA regions of $SwpA_{26\text{-}227\ TX1330}$, $SwpA_{26\text{-}226\ TX82}$, $LwpA_{33\text{-}772\ TX1330}$, $LwpA_{33\text{-}716\ TX82}$ $DufA_{45\text{-}309\ TX82}$, $SwpC_{26\text{-}243\ TX1330}$, $SwpC_{25\text{-}243\ TX82}$, $LwpC_{35\text{-}587\ TX1330}$ and $LwpC_{10\text{-}562\ TX82}$ were amplified using the primers found in Table 24 (all are recombinant proteins lacking the signal peptide, LwpA and LwpC lack the WxL domain, and DufA lacks the transmembrane domains). These fragments were cloned into the expression vector pQE30 (Qiagen) and transformed into M15 (pREP4) (Qiagen)

to obtain the plasmids and strains found in Table 20. The constructs were confirmed by sequencing. Recombinant proteins with N-terminal His_6-tags were then expressed by inoculating 4L of Luria Broth (LB) with 40mL of overnight culture of the respective expression strains. 1mM (for 4-6hr induction) 2mM (for overnight induction) IPTG was added to induce protein expression after cells reached mid-log phase (OD_{600}=0.6). Bacteria were harvested by centrifugation, the supernatant decanted, and the cell pellet washed and resuspended in 50mL of Buffer A (25mM Tris, 100mM NaCl, and 1mM DTT pH 7.9). The cells were then lysed using a French Press at 1300 psi and cell debris removed by ultracentrifugation at 40,000rpm for 30 minutes. The supernatant was then filtered through a 0.22µm membrane. The recombinant proteins were then purified using nickel affinity chromatography. A 5ml HisTrap chelating column was connected to an FPLC and charged with 100mM Ni^{2+} and equilibrated with Buffer A. The cell lysate was applied to the column and washed with 10 bed volumes of Buffer A, then the bound protein was eluted with a continuous linear gradient of imidazole in Buffer A (5-200mM; total volume 150mL) at a flow rate of 4.5mL/min. Fractions determined to have protein by absorbance at 280nm and SDS-PAGE were pooled, dialyzed into 25mM Tris and 1mM EDTA, pH 8.0, and applied to a 5mL HiTrap Q Column for anion exchange chromatography. Bound protein was eluted with a continuous linear gradient of NaCl (0-0.5M; total volume 200mL). Fractions containing purified recombinant protein were identified using SDS-PAGE, and dialyzed against PBS. Molecular masses were determined with MALDI-TOF MS. Predicted molecular weights, theoretical pI, and extinction coefficients were determined using the ExPASy ProtParam tool.

TABLE 24. Primers used to make WxL recombinant proteins.

Oligonucleotides[a]	5-3' Sequence[b]
TX1330	
WxL27PQE30F	CGCGGATCCGAAGGTCAAGCAACTTCAAA
WxL27PQE30R	GCGTCGACCTATGCTACTGGGCCAGCC
WxL16PQE30F	CGCGGATCCGTTGGAGCAACGCCTCCTAC
WxL16PQE30R	GCGTCGACACATCAAAAGAAAAATCGCC
WxL5A9PQE30F	CGCGGATCCGCTGAAGTATACCCAAAGGA
WxL5A9PQE30R	GCGTCGACCTAGTTATTGGGTGTATCTTC
WxL5B8PQE30F	CGCGGATCCTATTCTTTGGAATCGAATCAG
WxL5B8PQE30R	GCGTCGACATTTTCAATCCACCTGAGAAC
TX82	
pQE30LAWxL27TX82F	CGCGGATCCGAGGGTCAAGCGACTTCAAAAG
pQE30LAWxL27TX82R	GCGTCGACTGCTACTGGGCCAGCCACGAGGTTC
LApQE30DUFTX82F	CGCGGATCCCCAGAGAACCAAATAGATAAAG
LApQE30DUFTX82R	GCGTCGACTGGGAAAGATTGTTCTTGTACTG
pQE30LAWxL16TX82F	CGCGGATCCGTTGGAGCAACGCCTCCTGCC
pQE30LAWxL16TX82R	GCGTCGACGCCTTTTGCTTCATAATAACTTG
pQE30LCWxL5A9TX82F	CGCGGATCCTCTGCAACTGAAGTATACCCAAAG
pQE30LCWxL5A9TX82R	GCGTCGACGTTGTTAGGTGTATCTTCTAATG
WxL5B8TX82pQE30F	CGCGGATCCTACTCTTTGGAATTGAATCAG
WxL5B8TX82pQE30R	GCGTCGACATTTTCAACCCACCTGAGAAC

[a]WxL proteins were named differently when experiments first began. WxL27 is SwpA, WxL16 is LwpA, WxL410 is LwpB, WxL311 is SwpB, WxL5A9 is SwpC, WxL5B8 is LwpC.

[b] Enzyme restriction sites used for cloning are underlined

[4.2.6] Fluorescence and Circular Dichroism (CD). Steady-state intrinsic tryptophan fluorescence spectra were obtained on a Spectrofluorimeter LS 50B (Perkin-Elmer). Proteins were incubated in PBS (8 mM Na2HPO4, 2 mM KH2PO4, 3 mM KCl, 137 mM NaCl, pH7.4) or low salt PBS (4 mM Na2HPO4, 1 mM KH2PO4, 1.5 mM KCl, 68 mM NaCl, pH7.4) at RT for 1 hr before measurement. Samples were excited at 280 nm or 295 nm, and emission spectra

were collected, with an excitation slit of 5 nm and an emission slit of 5 nm. All spectra were corrected by subtraction of the appropriate blanks.

CD measurements in the far UV region (190-260 nm) were carried out at ambient temperature on a Jasco J-720 spectropolarimeter (Easton, MD) with a 0.5 mm cell. Proteins were dialyzed in phosphate buffer (0.8 mM Na2HPO4, 0.2 mM KH2PO4, 0.3 mM KCl, 13.7 mM NaCl, pH7.4)). Ten scans were collected and averaged at a scan speed of 100 nm/min, with a time constant of 2 s and band width of 1 nm. The spectra were background corrected with the CD signal obtained from the buffer. Deconvolution of the spectra was performed on DichroWeb server. Secondary structure composition for each protein are average of the deconvolution results from SELCON, CDSSTR, and CONTINLL using database 4.

[4.2.7] Production of polyclonal antibodies and purification of antigen-specific antibodies.

Production of antiserum was performed at previously described (Singh et al., 2012). Briefly, two Sprague-Dawley rats per antigen were immunized by intradermal injection three times over a six week period with 100µg of antigen in 200µL total volume of PBS and Freund's complete adjuvant (FCA) (1^{st} immunization) or Freund's incomplete adjuvant (FICA) (2^{nd} and 3^{rd} immunization). Rats were sacrificed and the anti-serum collected. Antisera were tested for antibody titers by ELISA (Nallapareddy et al., 2000). Briefly, 96-well plates were coated with 1µg of recombinant protein in 0.05M carbonate buffer, pH 9.6. Rat sera were tested in triplicate with serial dilutions from 1:50 to 1:128,000, followed by detection with AP-conjugated anti-rat secondary antibody and 1-step PNPP (Thermo Scientific). The reaction was stopped with 2M NaOH and absorbance read at 405nm. Antigen specific antibodies were then purified as described previously (Sillanpaa et al., 2008). Briefly, the corresponding antigen was coupled to CNBr-activated Sepharose 4B according to the manufacturer's instructions (GE Healthcare).

The antibodies were eluted from the column using 0.1M glycine, pH 2.8, neutralized with 1M Tris-HCl, pH8.0, and dialyzed against PBS. The antibodies were concentrated by ultrafiltration using the appropriate molecular mass cut off filter, and concentrations determined via absorption spectroscopy.

[4.2.8] Whole-cell ELISA for detection of WxL proteins on the bacterial cell surface. The presence of SwpA, LwpA, SwpC, and LwpC on the surface of bacterial cells was detected using affinity purified antigen specific antibodies in a whole-cell ELISA as previously described (Nallapareddy et al., 2003). Briefly, overnight cells of each bacterial strain were resuspended in carbonate buffer to an OD600=1.0, and coated in 96-well plates overnight at 4C. The cells were then washed with PBST and blocked with 2% BSA. A 1:200 dilution of each respective antibody in 100µl of 1% BSA was added to each well and incubated for 2hrs at 37C. Wells were then washed again with PBST, followed by the addition of 100µl of 1:2000 dilution of goat anti-rat IgG conjugated to HRP for 1hr at 37C. The wells were washed, TMB chromogenic solution added, then stopped with 2M sulfuric acid after 15 minutes, and absorbance read at 450nm. Antiserum against formalin-killed TX16 cells was used as a positive control to verify approximately equal binding of whole-cells to the microtiter plate wells. Statistical significance (*p*-values) of the surface exposure of each WxL proteins tested between the clinical and community groups of strains was calculated using an unpaired *t*-test.

[4.2.9] Mutanolysin cell-wall extraction and western blot analysis. Mutanolysin extracts were prepared as described previously (Gao et al., 2010; Nallapareddy et al., 2000). Briefly, TX82 cells were grown in BHI at 37C to late exponential phase and cell wall associated proteins extracted with mutanolysin (5U per each OD600=1 cells). Mutanolysin extracts were lyophilized to concentrate the extract, eluted in 200µl PBS, and protein concentrations

determined via BCA assay according to manufacturer's instructions (Pierce). Equal amounts of mutanolysin cell wall extracts were run on 4-15% gradient SDS-PAGE gels under reducing conditions, transferred to PVDF membranes, and probed using the specific antibodies to SwpA, LwpA, SwpC, and LwpC, followed by HRP-conjugated anti-rat IgG antibodies, and detected with Supersignal West Pico chemiluminescent substrate (Thermo Scientific). Antisera from pre-immune rats were used as a control.

[4.2.10] Immunogold electron microscopy. Immunogold electron microscopy was performed as described previously (Mandlik et al., 2008). Briefly, TX82 cells were grown on BHI agar overnight. The cells were harvested from the plates by scraping and then washed with 0.1M NaCl. For immunogold labeling, the bacterial suspension was spotted and fixed onto carbon grids, washed three times with 2% BSA in PBS, and blocked with 0.1% gelatin for 1hr. The samples were then incubated with the specific anti-WxL antibody (anti-SwpC, anti-LwpC, or pre-immune serum) in a 1:100 or 1:50 dilution for 1 hr, followed by washing and blocking as above. Then the samples were treated with 12-nm or 18-nm gold anti-rat IgG (1:20 dilution) in 2% BSA in PBS for 1 hr. The samples were then washed with water and stained with 1% uranyl acetate and viewed using a Jeol 1400 transmission electron microscope.

[4.2.11] ELISA-type ligand-binding assays and Surface Plasmon Resonance (SPR). Binding of recombinant His-tagged proteins from Locus A to ECM proteins was tested as previously described (Nallapareddy et al., 2003). Briefly, Immunoblon 4HBX 96-well plates were coated with 1µg of ECM protein in 100µl of PBS at 4C overnight. After washing with PBST and blocking with 2% BSA, wells were incubated with 20µM of His-tagged recombinant protein at 37C for 4 hours. After washing unbound recombinant protein with PBST, the bound His-tagged recombinant proteins were detected with mouse anti-His followed by AP-

conjugated anti-mouse antibody and detected with 1-step PNPP according to manufacturer's instructions (Thermo Scientific). The reaction was stopped with 2M NaOH and the absorbance read at 405nm.

SPR-based Biacore binding experiments were performed at 25^{o} C on a Biacore 3000 (GE Healthcare). Fibronectin and Hep2-40K (Millipore) were immobilized on CM3 sensor chip, and type I collagen, DufA, SwpA, and LwpA were immobilized on CM5 sensor chip using amine coupling procedure as recommended by the manufacturer. A reference surface was prepared with activation and deactivation treatments but no protein coupled. PBS including 0.005% Tween 20 was used as running buffer. Soluble proteins at different concentrations were injected onto the immobilized ligand surface to obtain SPR response curves (sensorgrams).

Baseline corrected sensorgrams (with buffer blank run further subtracted) were globally fitted to two-state model using BIAevaluation software (Version 4.1). The association rate constant (k_{a1}) and dissociation rate constant (k_{d1}) for the binding state, forward rate constant (k_{a2}) and backward rate constant (k_{d2}) for the conformational change state were derived from global fitting. The apparent dissociation constant (K_D^{app}) was calculated according to the formula $K_D = 1 / ((k_{a1} / k_{d1}) * (1+k_{a2} / k_{d2}))$.

[4.2.12] Single cross-over insertion mutants and non-polar markerless deletion mutants.

All mutants were created using an altered PheS* system as described previously (Kristich et al., 2007; Panesso et al., 2011). For single-cross over insertion mutants, small fragments of the first gene of each operon of TX1330 flanked by rho-independent terminators were amplified by PCR using the primers in Table 25, and cloned into pHOU1. The correct insert was confirmed by sequencing using plasmid specific primers AB142 and AB143 (Table 25) and the plasmid

electroporated into *E. faecalis* CK111. Blue colonies were recovered on BHI Gentamicin 125µg/mL, X-Gal 200µg/mL plates. These recombinant plasmids were then introduced into TX1330RF through filter mating with CK111 as the donor. Single-cross over integrants were then selected on BHI Gentamicin 125µg/mL, Rifampin 100µg/mL, Fusidic Acid 25µg/mL plates and confirmed using the primers AB142 and the reverse primer specific to the Locus, as well as AB143 and the forward primer specific to the Locus (Table 25).

TABLE 25. Primers used for single cross-over insertion mutants in TX1330.

Oligonucleotide	**5-3' Sequence[a,b]**
Locus A	
271XoverF	TTGCGGCCGC**AAAAAAGCGGCCAAAGGCCTGCCACCAGGCC AAAGGCCGCTTTTTT**ACTTCTTCTAGGCGGCT
271XoverR	CCGGAATTC**AAAAAAATTGAAGCAAGTGTTTCCACCATTGCT TCAATTTTTT**ACACCTCGACGATCTGAT
Locus B	
4101XoverF	TTGCGGCCGC**AAAAAAGCGGCCAAAGGCCTGCCACCAGGCC AAAGGCCGCTTTTTT***CGTGCAACTGTCACATTAG*
4101XoverR	CCGGAATTC**AAAAAAATTGAAGCAAGTGTTTCCACCATTGCT TCAATTTTTT**ATTCCGCGATTGTTGTC
Locus C	
5A91XoverF	TTGCGGCCGC**AAAAAAGCGGCCAAAGGCCTGCCACCAGGCC A<u>AAGGCCGCTTTTTT</u>**TTGAAGCAGGAGATGAGG
5A91XoverR	GCGGGATCC**AAAAAAATTGAAGCAAGTGTTTCCACCATTGCT TCAATTTTTTT**CTGTGCTCCAACTAGTTC
To confirm sequence/ plasmid insertion	
AB142	AAGAAAGCCTACCCTTGG
AB143	CACGAACGAAAATCAAGC

[a] Restriction enzyme sites for cloning are underlines

[b] Terminator sequences are in bold

TABLE 26. Primers used for markerless non-polar deletions on the WxL Loci in TX82.

Oligonucleotide	5-3' Sequence
Locus A Deletion in TX82	
LocusA6geneDelUpF1	TTGCGGCCGCCGTATTGACAGCGGATGCAACAG
LocusA6geneDelUpR2	ACTCTATTGCATGTCCCAATCATGTTCTTTTTCAT
LocusA6geneDelDownF3	ATTGGGACATGCAATAGAGTAGATAAAAGACTGTGA
LocusA6geneDelDownR4	GCGGGATCCACCGCATCAGGTAGAACAAG
LAUDInternalF	GCTGGAACACAAACAGTAGAAG
LAUDInternalR	CTTAGTAATCATGGACATCCCTAG
LAOutsideUpF	GAAGAGATAAGAGAAGCAGGC
LAOutsideDownR	CTAACACAGCTACTACCAGCATC
Locus B Deletion in TX82	
LocusBDelUpF1	TTGCGGCCGCCGTATTGACAGCGGATGCAACAG
LocusBDelUpR2	TGATATGATTTCTCCACCATCCTCTCTGACGTTTT
LocusBDelDownF3	ATGGTGGAGAAATCATATCAACGATTTCAGGGAGT
LocusBDelDownR4	GCGGGATCCCAGGTATTGATGTCGCCGAACAAAT
LBUFInternalF	GCTGGAACACAAACAGTAGAAG
LBDFInternalR	CTTAGTAATCATGGACATCCCTAG
LAOutsideUpF	GAAGAGATAAGAGAAGCAGGC
LAOutsideDownR	CTAACACAGCTACTACCAGCATC
Locus C Deletion in TX82	
LocusCDelupF9	TTGCGGCCGCGATCGGATGATTTGTAACTCG
LocusCDelupR10	CTCTTTTCCGATTTCTTTTTTCATCATTGACATTCC
LocusCDelDownF11	GAAATCGGAAAAGAGAAAGTTGACGAGAGTATAA
LocusCDelDownR12	GCGGGATCCCTTTAGTGACTAGAGTTTGCAC
LocusCOutsideUpF	AATCAGCGAATGTTTATTT
LocusCOutsideDownR	CTTTAGTGACTAGAGTTTGCAC
Locus C Internal F	GTACTAGATGCAAGCGTGTATTC
Locus C Internal R	CATTCATAGGGAAGTCCTCTTTTC

For deletion of WxL loci, the deletion construct fragments (upstream and downstream fragments amplified together by cross-over PCR) amplified using the primers found in Table 26 and cloned into pHOU1. The correct insert was confirmed by sequencing and the plasmid electroporated into *E. faecalis* CK111. Blue colonies were recovered on BHI Gentamicin 125µg/mL, X-Gal 200µg/mL plates. These recombinant plasmids were then introduced into TX82 through filter mating with CK111 as the donor. Single cross-over integrants were then

selected on BHI Gentamicin 125µg/mL, Vancomycin 50µg/mL or BHI Gentamicin 125µg/mL, Erythromycin 200µg/mL plates, replica plated on BHI Gentamicin 125µg/mL, Ampicillin 125µg/mL plates, and confirmed using the primers in Table 26. Positive colonies were then streaked onto MM9YEG 10mM p-Chl-Phe plates and incubated at 37C. Colonies that grew on p-Chl-phe were then replica plated on BHI Gent125µg/mL to confirm excision of the plasmid. Colonies susceptible to gentamicin were then screened by PCR for the deletion (Table 26), sequenced, and PFGE performed to confirm the correct parental background.

[4.2.13] Whole-cell ligand binding assay, biofilm assay, and bile-salt assay. Whole-cell binding of TX82 and WxL locus mutants were tested by an ELISA-like binding assay as described previously (Pinkston et al., 2011). Immunoblon H4BX 96-well plates were coated with 1µg of ECM protein in 100µl of PBS overnight at 4C. Wells were blocked with 2% BSA 0.1% Tween20 in PBS. *E. faecium* cells were cultured in BHI overnight at 37°C, then the overnight bacterial cultures were diluted in BHI to an OD_{600} of 0.05 and grown to mid-log phase (~ 4 hours) at 37°C. The cells were pelleted, washed twice with PBS, and resuspended in 1%BSA 0.05% Tween20 to an OD600=1. 100µl of cells were added to each well and incubated for 2hrs at RT. Non-bound cells were removed by washing with PBS, and the remaining cells were fixed using 100µl/well of Bouin's fixative for 30 minutes. The wells were washed and bound cells stained with 1% crystal violet 100µl/well for 30 minutes. Extraneous dye was then washed away with water, and then solubilized using 100µl of ethanol/acetone (80:20) for 20 minutes. The absorbance was read at 570nm. To compare the mean values of the adherence results, an unpaired *t*-test was used.

Biofilm formation by TX82 and WxL locus mutants was assayed as described previously (Mohamed et al., 2004). Briefly, cells were grown overnight in 10 mL of TSBG (all

normalized to an OD600=1 with the addition of fresh TSBG). The cells were then diluted 1:100 in TSBG and 200µl of each dilution added to 96-well plates. After incubation for 24hrs at 37C, unbound cells were removed by washing with PBS, followed by adding 200µl/well of Bouin's fixative for 30 minutes. The wells were washed and biofilm cells stained with 1% crystal violet 200µl/well for 30 minutes. The extraneous dye was then washed away with water, and then solubilized using 200µl of ethanol/acetone (80:20) for 20 minutes. The absorbance was read at 570nm.

Resistance to bile salts was determined using the protocol described previously (Teng et al., 2005). Briefly, cells were grown for 24hrs at 37C in BHI. The cells were then pelleted, washed with PBS, and resuspended in the same volume of fresh BHI. 20µl of the overnight suspension was then added to 1ml of BHI containing 0.3% bile salts. After 30 minutes of incubation at 37C, *E. faecium* cells were serially diluted and plated on BHI agar for colony forming units (CFU) counts. CFUs were counted after 24hrs and the percent survival was calculated by comparing CFUs after 30 minutes to CFUs at time zero using the equation previously described (Teng et al., 2005). Percent survival= $\{[(\text{CFU of mutant})_{t=30})/(\text{CFU of mutant})_{t=}0]\}/ \{[(\text{CFU of TX82})_{t=30})/(\text{CFU of TX82})_{t=}0]\}$ multiplied by 100. Wild-type and mutant cells were then compared in their survival percentage of wild-type after bile salt exposure. To compare the mean survival values of the mutant versus TX82, an unpaired *t*-test was used.

[4.2.14] Detection of antibodies to WxL proteins in patient sera and immunization/ protection studies. Detection of antibodies in patient sera was performed at described previously (Nallapareddy et al., 2008b). Briefly, recombinant SwpA, LwpA, SwpC, and LwpC were run on 12% SDS-PAGE gels and transferred to PVDF membranes. The membranes were

blocked with 2% milk and 0.1% Tween20 and then incubated with either endocarditis patient serum or healthy volunteer serum (1:2000) at RT for 2hrs. After washing with PBST, antibodies were detected with goat anti-human IgG (1:5000) conjugated to HRP at RT for 1hr. The western was developed using Supersignal West Pico chemiluminescent substrate (Thermo Scientific).

Active immunization and challenge with bacterial endocarditis was performed as previously described (Singh et al., 2012). Briefly, Sprague Dawley rats were immunized intradermally with either 1) a mix of recombinant LwpA and LwpC in FCA or FICA 2) sham immunization (FCA and FICA) or 3) not immunized, creating three groups. Group 1 received an initial dose of 100µg of recombinant LwpA and 100µg of recombinant LwpC in Fruend's complete adjuvant (FCA), followed by two boosters at two week intervals of the same dose of LwpA and LwpC in Fruend's incomplete adjuvant (FICA). Sham immunizations were performed without recombinant protein in the same manner. Rats were then challenged by creating sterile vegetations in the heart valves of the three groups of rats. Rats were induced with experimental endocarditis and catheterized as described previously (Singh et al., 2012), then 24hrs post catheterization, inoculated via i.v. tail vein injection with 10^9 cells of TX82. The animals were then sacrificed 48hrs post infection and the vegetations harvested and processed as described previously (Singh et al., 2012). CFUs in the vegetations were then determined between the three groups. Differences in the bacterial $\log_{10}$ CFU (geometric mean) of the heart vegetations from the three groups were analyzed by unpaired *t*-test. The Fisher's exact test was then used to compare the total number of infected vs. non-infected in the three groups. Graph Pad Prism 4.0 was used to determine the statistical values.

[4.3] CHAPTER 3 RESULTS

As discussed in the previous chapters, hospital outbreaks and the global dissemination of multi-resistant *E. faecium* have been attributed to a hospital-associated clade (Galloway-Pena et al., 2012). HA-clade isolates contain higher rates of putative virulence determinants encoding surface-associated proteins (Galloway-Pena et al., 2009; Sillanpaa et al., 2009a). Among these are genes encoding MSCRAMM (microbial surface components recognizing adhesive matrix molecules)-like proteins harboring domains which include a secretion signal sequence, ligand-binding domains with Ig-like folds, and a C-terminal cell-wall anchoring domain with an LPxTG-like motif (Hendrickx et al., 2009b). Surface protein-mediated attachment to host tissues and subsequent colonization is a requirement for pathogenesis (Hendrickx et al., 2009b). Therefore, proteins of this nature, or proteins within the same operon, could contribute to a selective advantage in the nosocomial environment by increasing the colonizing ability of strains (Galloway-Pena et al., 2009; Hendrickx et al., 2009b; Nallapareddy et al., 2008b; Nallapareddy et al., 2003; Sillanpaa et al., 2008; Sillanpaa et al., 2009a).

Proteins containing the newly described WxL-domain in their C-termini were recently identified in 17 low G+C content gram positive species, including the close relative of *E. faecium*, *E. faecalis* (Bourgogne et al., 2008; Brinster et al., 2007a; Siezen et al., 2006). A recent study demonstrated that the C-terminal region of two WxL proteins in *E. faecalis* conferred surface localization and that the WxL domains are necessary for interactions with the cell wall. This study also showed that these WxL domains could bind the cell-wall fraction of six other species of gram-positive bacteria and the purified peptidoglycan of *Bacillus subtilis*. Furthermore, when the WxL domain was deleted, the interaction with other bacterial species

and purified peptidoglycan was abolished. Therefore, the WxL domain has been proposed as a novel cell-wall binding domain (Brinster et al., 2007a).

Previous studies have found that some WxL proteins exist in clusters in which most proteins encoded within the cluster contain a signal peptide for secretion by the Sec-dependent pathway and most clusters also contain genes encoding the following, i) a small protein with a cell-wall binding LPxTG motif, ii) a small WxL protein with Trp-x-Leu, Asp-x-Arg-Gly, and proline rich domains, iii) a large WxL protein with possible concanavalin A-like lectin domains and iv) a DUF916 family membrane protein with a C-terminal transmembrane helix (Brinster et al., 2007a; Siezen et al., 2006). Due to the fact that these WxL genes have been proposed to encode novel cell-wall binding domains and exist in tandem with LPxTG proteins, it seems likely there is some form of assembly of these proteins on the cell surface. Although previous studies have elucidated that these genes are co-transcribed (Brinster et al., 2007a; Siezen et al., 2006), no studies have examined the possible interactions between the proteins encoded by the genes within the putative operon.

Interestingly, while we were looking for putative MSCRAMMS in the *E. faecium* strain TX16 as discussed in Chapter 2, we noticed some genes encoding atypical MSCRAMMs, one of these had been named *fms6* and was found in a locus with putative WxL proteins (Locus A). The gene *fms6* encodes a protein that contains typical MSCRAMM motifs and domains such as a signal sequence, an LPxTG-like motif (FPETN), a membrane spanning region, and C-terminal positive residues. However, the predicted Fms6 protein is only 116 amino acids, which is much smaller than other studied MSCRAMMs that are usually greater than 650 amino acids. Due to the fact that the WxL domain was suggested to be a novel cell-wall binding domain and

that two of these were found in a locus with an atypical MSCRAMM, we decided to look for more of these proteins within the *E. faecium* genomes.

[4.3.1] Six genes encoding WxL proteins in three different loci are found in *E. faecium* genomes. Our group recently sequenced three *E. faecium* genomes, two clinical endocarditis isolates TX16 (DO) and TX82, and a community fecal isolate, TX1330. Upon sequencing these three genomes, six genes encoding proteins with the WxL domain were found in each genome (Figure 15). *In silico* analysis showed that there are three clusters in each genome resembling the pattern found in *E. faecalis* and *L. plantarum* (Brinster et al., 2007a; Siezen et al., 2006) flanked by predicted rho-independent terminators (Figure 16). All three loci (named WxL Locus A, B, and C) contain genes that encode i) a small WxL protein (SwpA-C) with Trp-x-Leu, Asp-x-Arg-Gly, and proline rich domains, ii) a large WxL protein (LwpA-C) with possible concanavalin A-like lectin domains and, iii) a DUF916 family protein (DufA-C) with 2 transmembrane helices. Locus A (named arbitrarily) contains an LPxTG-like protein (*fms6*), but the other two loci do not contain a definitive LPxTG-like protein. Locus A and Locus C both have Mga-like transcriptional regulators associated with the loci, whereas Locus B has a LytR-family transcriptional regulator directly upstream of the predicted operon.

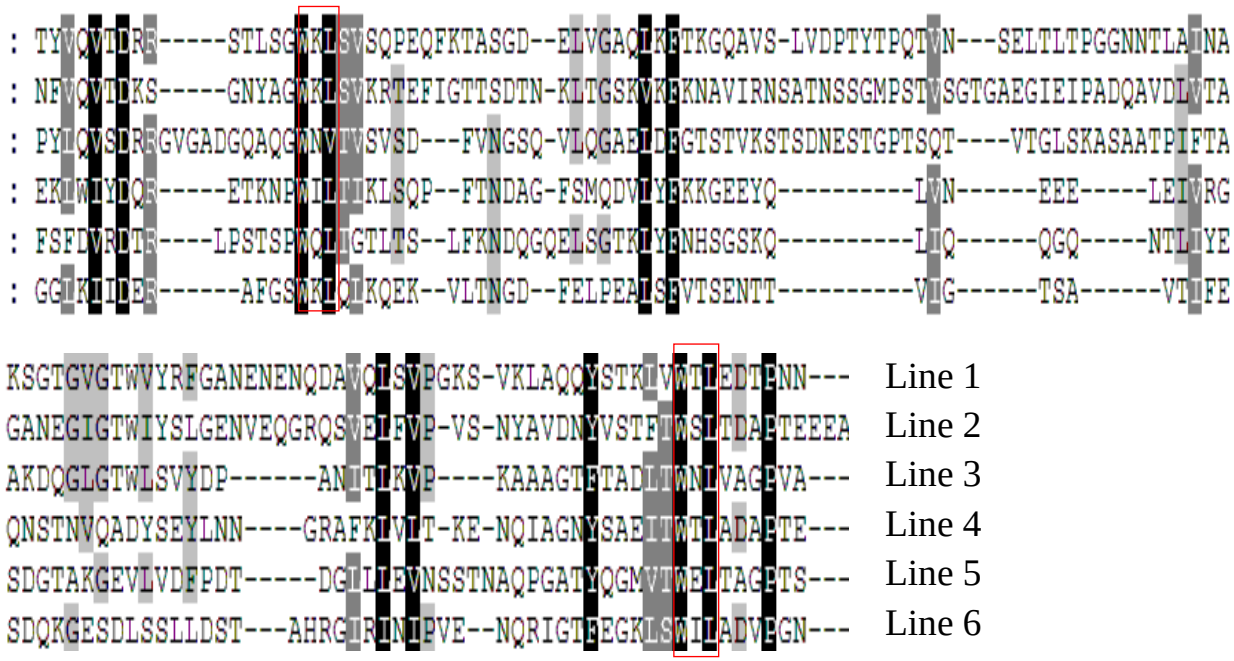


FIGURE 15. The WxL domain of 6 WxL proteins in *E. faecium*. An alignment of the C-terminus of the 6 WxL proteins found in *E. faecium*. Highly conserved residues are in black, moderately conserved residues are in grey, and the red boxes highlight the 2 "WxL" motifs. Lines 1) SwpC 2) SwpB 3) SwpA 4) LwpB 5) LwpA 6)LwpC.

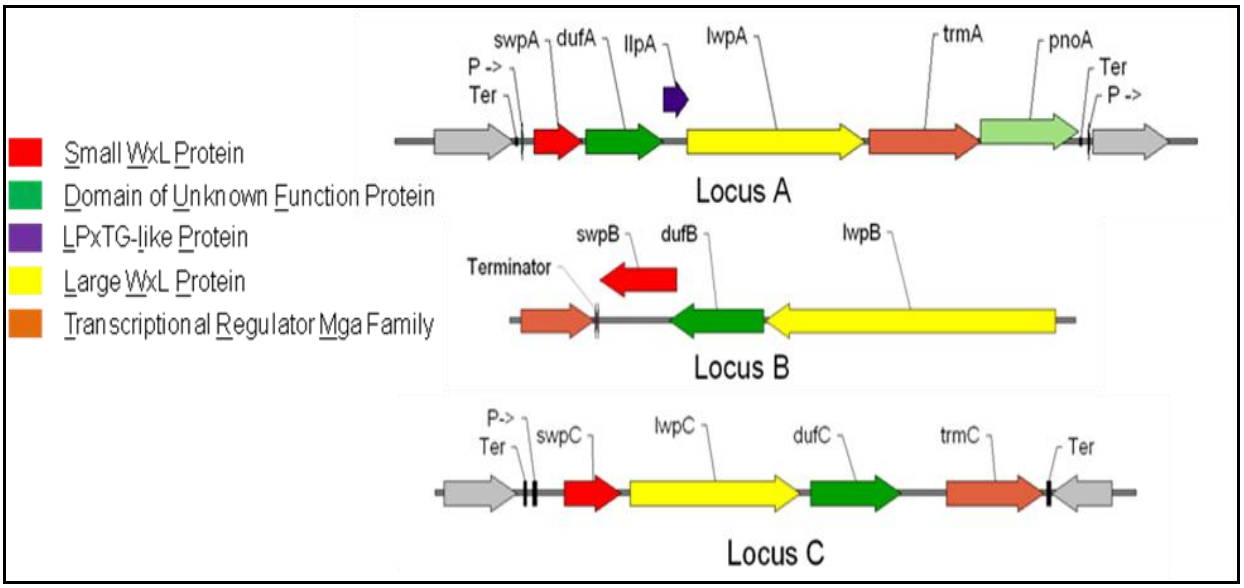


FIGURE 16. The organization of the three putative operons containing WxL genes in *Enterococcus faecium*. A pictorial representation of the 3 WxL putative operons found in *E. faecium* based on *in silico* analysis. Each arrow represents a predicted ORF named for the protein it encodes. Predicted terminators and promoters are also labeled.

RT-PCR of the individual WxL genes and the intergenic regions was performed on TX1330 and TX82 to confirm these genes are expressed and co-transcribed and therefore most likely in an operon. These results can be found in Figure 17 and 18.

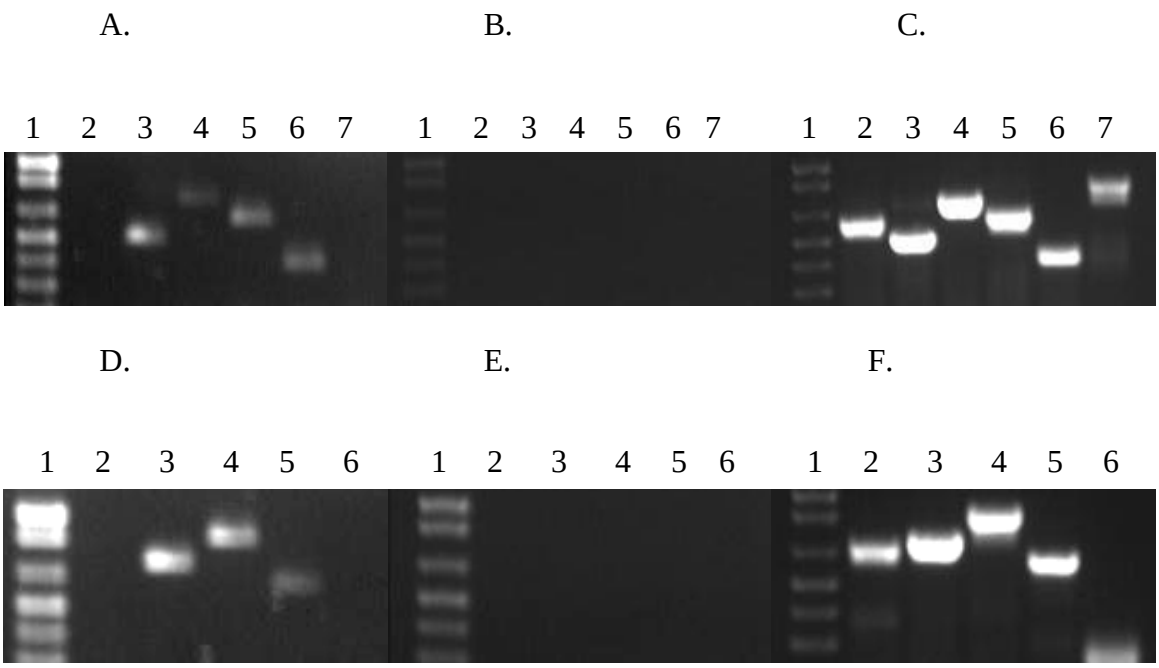


FIGURE 17. RT-PCR of intergenic regions of Locus A and C in TX1330 (CA-clade strain). Panels A through C show the RT-PCR analysis of the WxL Locus A of TX1330. Lanes: 1, 1KB Plus DNA Ladder; 2, OG (outside gene) to *swp;* 3, *swpA* to *dufA*; 4, *dufA* to *lwpA*; 5, *lwpA* to *trmA*; 6, *trmA* to *pnoA*; 7, *pnoA* to OG2. Panel A represents the reactions containing the reverse transcriptase and RNA, Panel B represents the reactions containing Taq polymerase and RNA, and Panel C represents the reactions containing Taq polymerase and genomic DNA. Panels D through F show the RT-PCR analysis of the WxL Locus C of TX1330. Lanes: 1, 1KB Plus DNA Ladder; 2, OG to *swpC* 3; 3, *swpC* to *lwpC*; 4, *lwpC* to *dufC*; 5, *dufC* to *lwpC*; 6, *lwpC* to OG2. Panel D represents the reactions containing the reverse transcriptase and RNA, Panel E represents the reactions containing Taq polymerase and RNA, and Panel F represents the reactions containing Taq polymerase and genomic DNA.

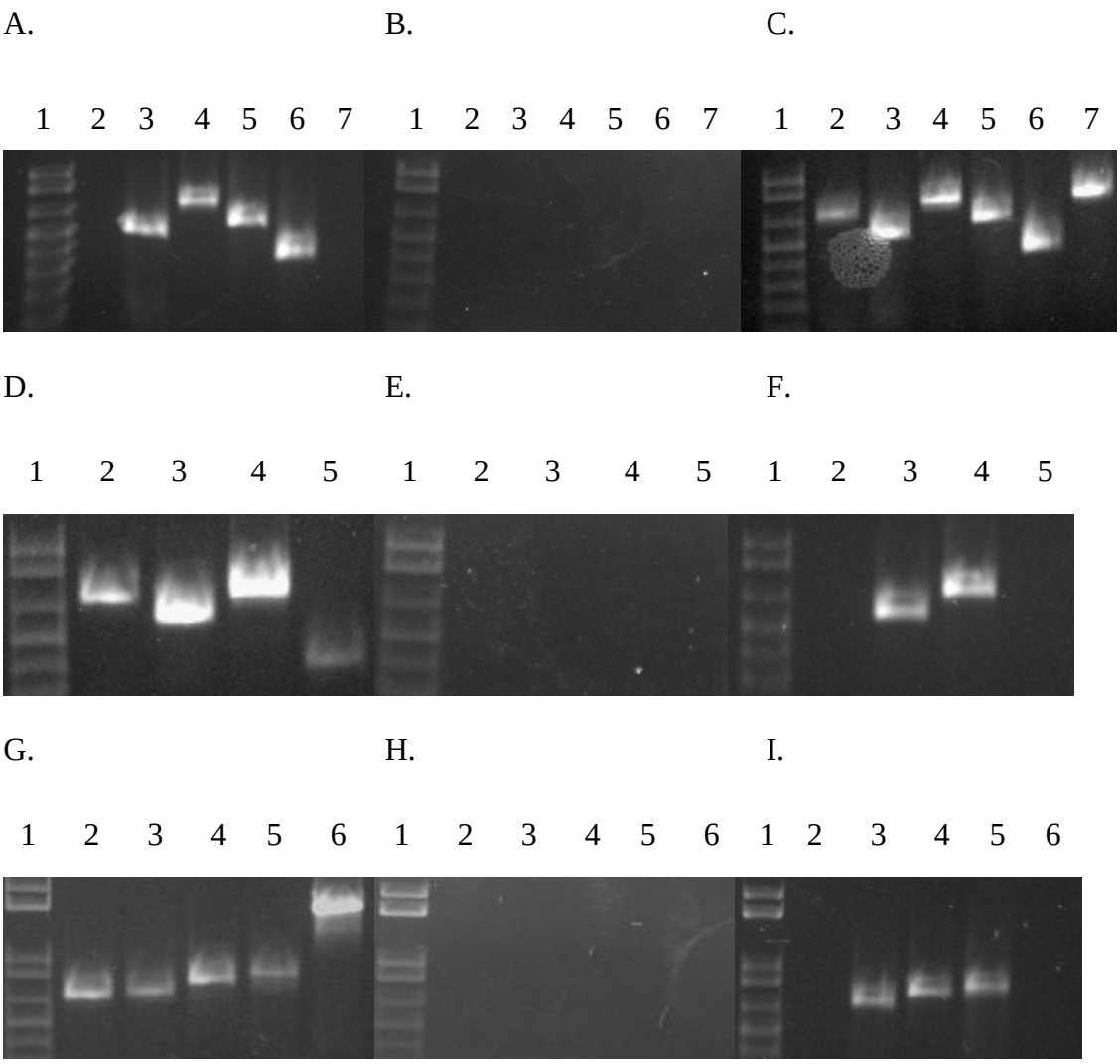


FIGURE 18. RT-PCR of intergenic regions of Locus A, B, and C in TX82 (HA-clade strain). Panels A through C show the RT-PCR analysis of the WxL Locus A of TX82. Lanes: 1, 1KB Plus DNA Ladder; 2, OG to *swpA*; 3, *swpA* to *dufA*; 4, *dufA* to *lwpA*; 5, *lwpA* to *trmA*; 6, *trmA* to *pnoA*; 7, *pnoA* to OG2. Panel A represents the reactions containing reverse transcriptase and RNA, Panel B represents the reactions containing Taq polymerase and RNA, and Panel C represents the reactions containing Taq polymerase and genomic DNA. Panels D through F show the RT-PCR analysis of the WxL Locus B of TX82. Lanes: 1, 1KB Plus DNA Ladder; 2, OG1 to *lwpB*; 3, *lwpB* to *dufB*; 4, *dufB* to *swpB*; 5, *swpB* to OG2. Panel D represents the reactions containing the reverse transcriptase and RNA, Panel E represents the reactions

containing Taq polymerase and RNA, and Panel F represents the reactions containing Taq polymerase and genomic DNA. Panels G through I show the RT-PCR analysis of the WxL Locus C of TX82. Lanes: 1, 1KB Plus DNA Ladder; 2, OG1 to *swpC*; 3, *swpC* to *lwpC*; 4, *lwpC* to *dufC*; 5, *dufC* to *trmC*; 6, *trmC* to OG2. Panel G represents the reactions containing the reverse transcriptase and RNA, Panel H represents the reactions containing Taq polymerase and RNA, and Panel I represents the reactions containing Taq polymerase and genomic DNA.

[4.3.2] The phylogenetics and prevalence of WxL genes throughout *E. faecium* genomes.

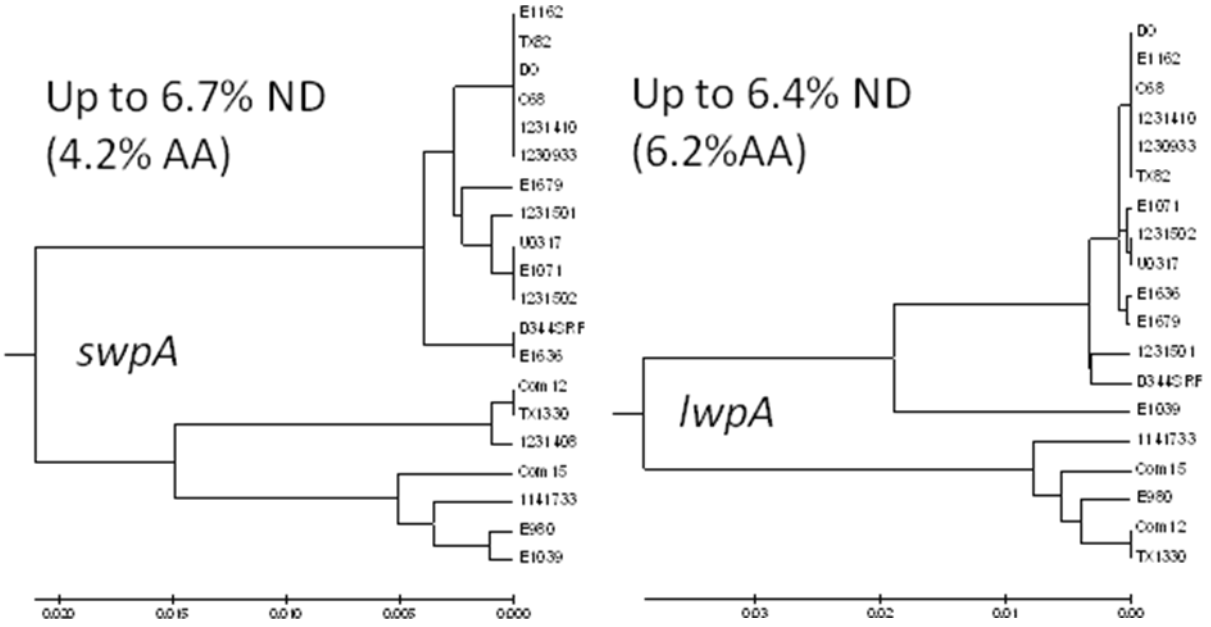
Further analysis of the distribution and homology of the genes in the WxL loci of the other sequenced *E. faecium* genomes (Chapter 2) found that all the genes within the three loci have homologs. Most of the homology ranges from ~86-100% nucleotide sequence identity, and 86-100% amino acid identity. The absence of homologs in some strains may be explained by the fact that they are incomplete genomes. Because we had found that *swpC* (named *wlcA* in previous paper) showed cladal separation (Chapter 1 and (Galloway-Pena et al., 2011)), we tested the other genes encoding WxL proteins for their phylogeny (Figure 19).

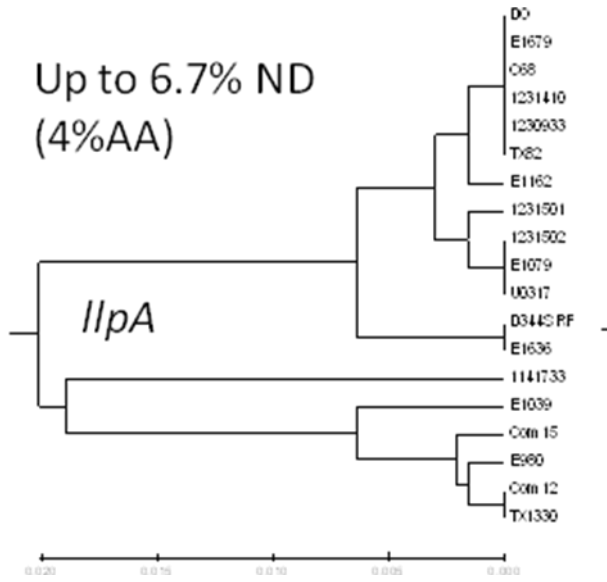
Interestingly, the predicted genes found upstream and downstream of the predicted operons varied from strain to strain, but had some consistency within strains of the same clade. These results can be found in Figure 20. In all strains analyzed, upstream of Locus A is a putative FMN-dependent alpha hydroxyl acid dehydrogenase and downstream is a putative ABC transporter. Upstream of Locus B, all the HA clade strains and the CA clade strains TX1330 and Com12, have a putative diaglycerol kinase directly, whereas in 1,231,408 and Com15, there is a transposon between the WxL Locus B and the diglycerol kinase gene. 1,141,733 is different in that it has three hypothetical proteins between the putative diaglycerol kinase and the WxL Locus B and lacks the transposon. Downstream of WxL Locus B, all the HA clade strains, and two of the CA clade strains (TX1330 and Com12), have a putative histidine kinase, a conserved hypothetical protein, and a sortase (corresponding to SrtA2 in TX82), whereas in the CA clade isolates 1,141,733 and Com15, and the hybrid isolate 1,231,408, have a putative ABC-Transporter ATP-binding protein. For WxL Locus C, HA and CA clade isolates differ as to what is upstream, both have a major intrinsic protein downstream, as well as other genes downstream that are the same, HA clade isolates have a putative

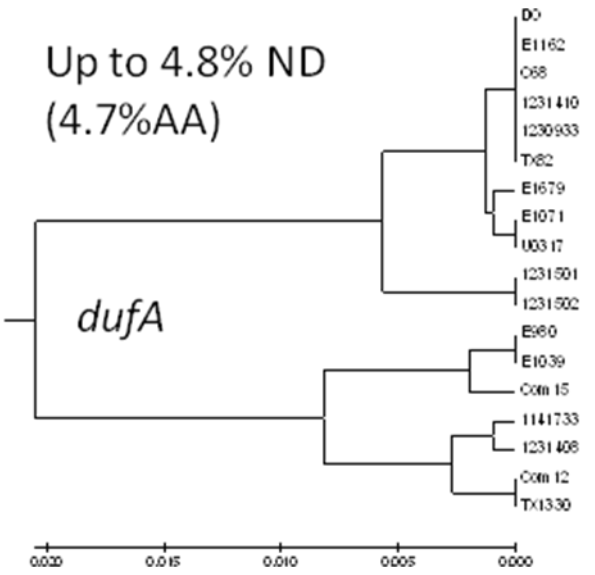
metallophosphoesterase upstream (and other genes that are different) whereas the CA clade isolates have a tagatose-1,6-diphosphate aldolase and other genes that are different from the clinical clade upstream of WxL Locus C.

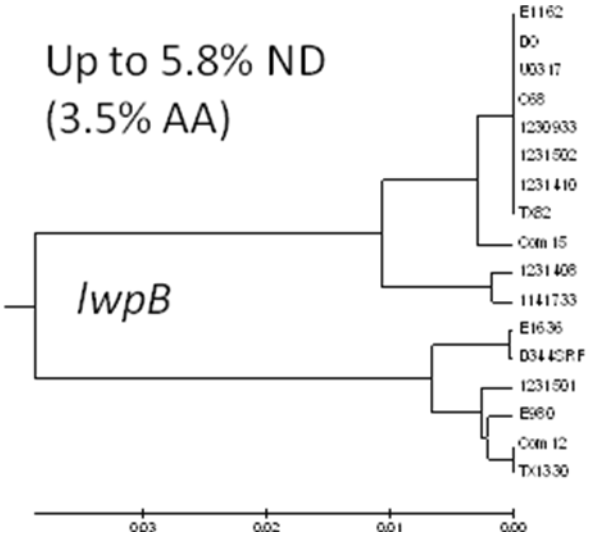
In addition to looking at sequenced genomes, colony hybridizations were performed under high stringency conditions with probes for each individual WxL gene using 54 clinical, 10 fecal, and 20 endocarditis isolates from our collection. All of the WxL genes were found to be present in all of the strains except for the absence of *lwpB*, *swpB*, and *lwpC* from several strains, which could possibly be explained by specificity of the probe (Figure 21).

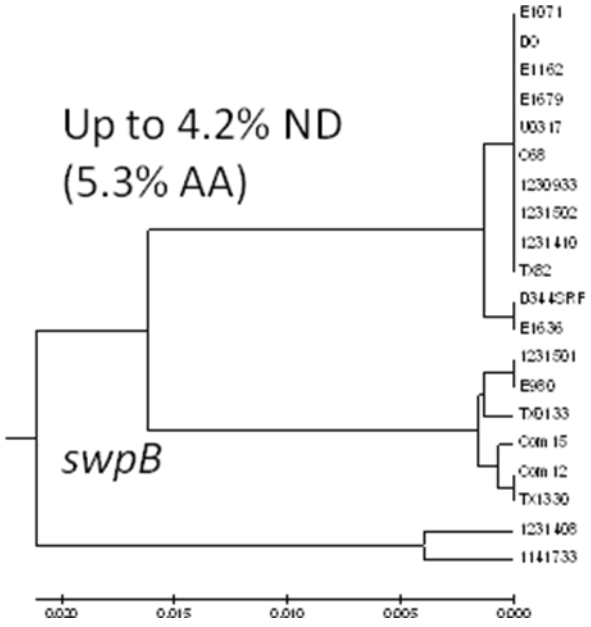
FIGURE 19. UPGMA phylogenetic trees and identities of WxL loci genes throughout the *E. faecium* genomes. The results are based on a pair-wise analysis of each gene's nucleotide sequence for each of the 21 *E. faecium* strains using the Poisson correction method in MEGA4.0.2 software. An UPGMA tree was constructed depicting the evolutionary distance between these sequences. The tree is drawn to scale with the branch lengths representing the evolutionary distances, the scale can be seen at the bottom of the tree. The largest nucleotide and amino acid differences between any two alleles are noted. ND denotes nucleotide difference. AA denotes amino acid.

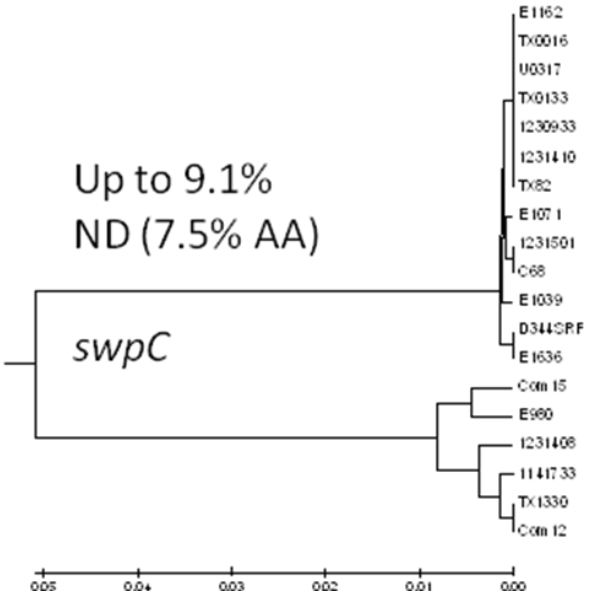


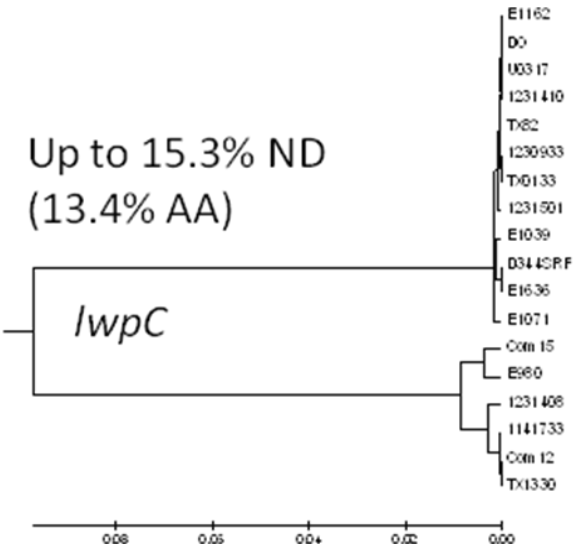

Up to 6.7% ND
(4%AA)
llpA
D0
E1679
O68
1231410
1230933
TX82
E1162
1231501
1231502
E1071
U0317
D344SPF
E1636
1141733
E1039
Com 15
E960
Com 12
TX1330
0.020
0.015
0.010
0.005
0.000

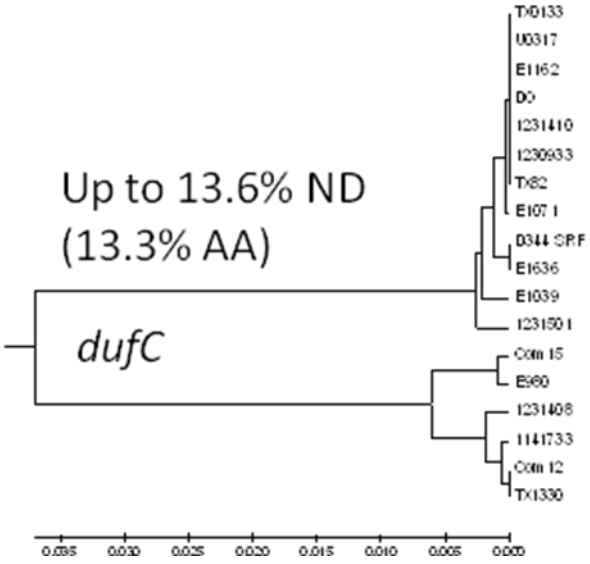

Up to 4.8% ND
(4.7%AA)
dufA
D0
E1162
O68
1231410
1230933
TX82
E1679
E1071
U0317
1231501
1231502
E960
E1039
Com 15
1141733
1231408
Com 12
TX1330
0.020
0.015
0.010
0.005
0.000


Up to 5.8% ND
(3.5% AA)
lwpB
E1162
D0
U0317
O68
1230933
1231502
1231410
TX82
Com 15
1231408
1141733
E1636
D344SPF
1231501
E960
Com 12
TX1330
0.03
0.02
0.01
0.00


Up to 4.2% ND
(5.3% AA)
swpB
E1071
D0
E1162
E1679
U0317
O68
1230933
1231502
1231410
TX82
D344SPF
E1636
1231501
E960
TX0133
Com 15
Com 12
TX1330
1231408
1141733
0.020
0.015
0.010
0.005
0.000


Up to 9.1%
ND (7.5% AA)
swpC
E1162
TX0016
U0317
TX0133
1230933
1231410
TX82
E1071
1231501
C68
E1039
D344SPF
E1636
Com 15
E980
1231408
1141733
TX1330
Com 12
0.05
0.04
0.03
0.02
0.01
0.00


Up to 15.3% ND
(13.4% AA)
lwpC
E1162
DO
U0317
1231410
TX82
1230933
TX0133
1231501
E1039
D344SPF
E1636
E1071
Com 15
E980
1231408
1141733
Com 12
TX1330
0.08
0.06
0.04
0.02
0.00


Up to 13.6% ND
(13.3% AA)
dufC
TX0133
U0317
E1162
DO
1231410
1230933
TX82
E1071
D344SPF
E1636
E1039
1231501
Com 15
E980
1231408
1141733
Com 12
TX1330
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000

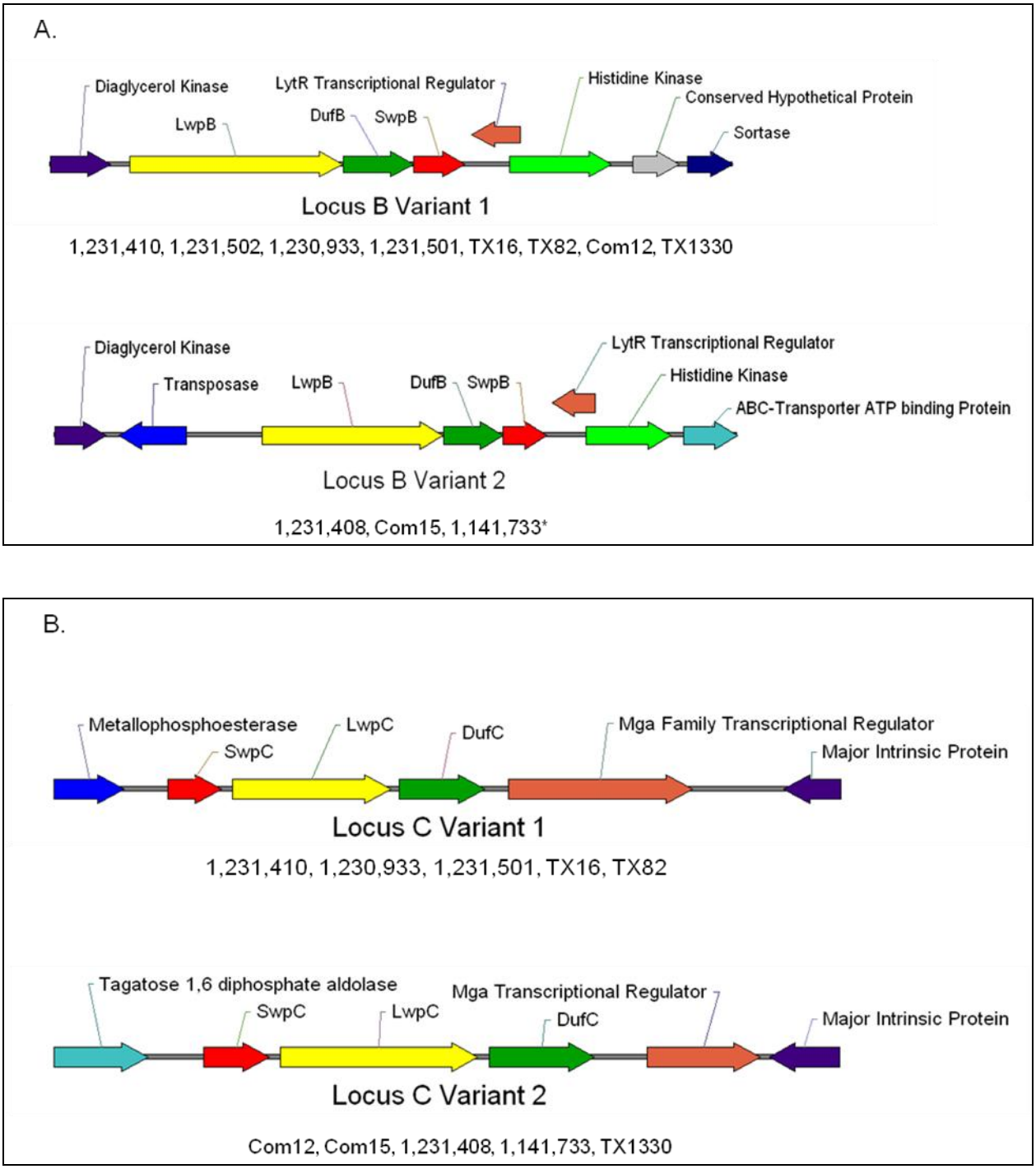


FIGURE 20. Variations upstream and downstream of the WxL loci. The genes upstream and downstream of the WxL loci of all Broad Institute strains and the strains sequenced by our

laboratory were analyzed. The genes upstream and downstream of WxL Locus A encoded the same predicted proteins in all of the strains. Panel A depicts the variants of the WxL Locus B surrounding region. The asterisk next to strain 1,141,733 denotes that it lacks the transposase and has 3 hypothetical proteins between the WxL Locus B and the putative diaglycerol kinase. Otherwise it has the same genes upstream and downstream as 1,231,408 and Com 15. Panel B depicts the variants of the WxL Locus C surrounding region. Strain 1,231,502 did not have a complete sequence in this region, and therefore could not be analyzed.

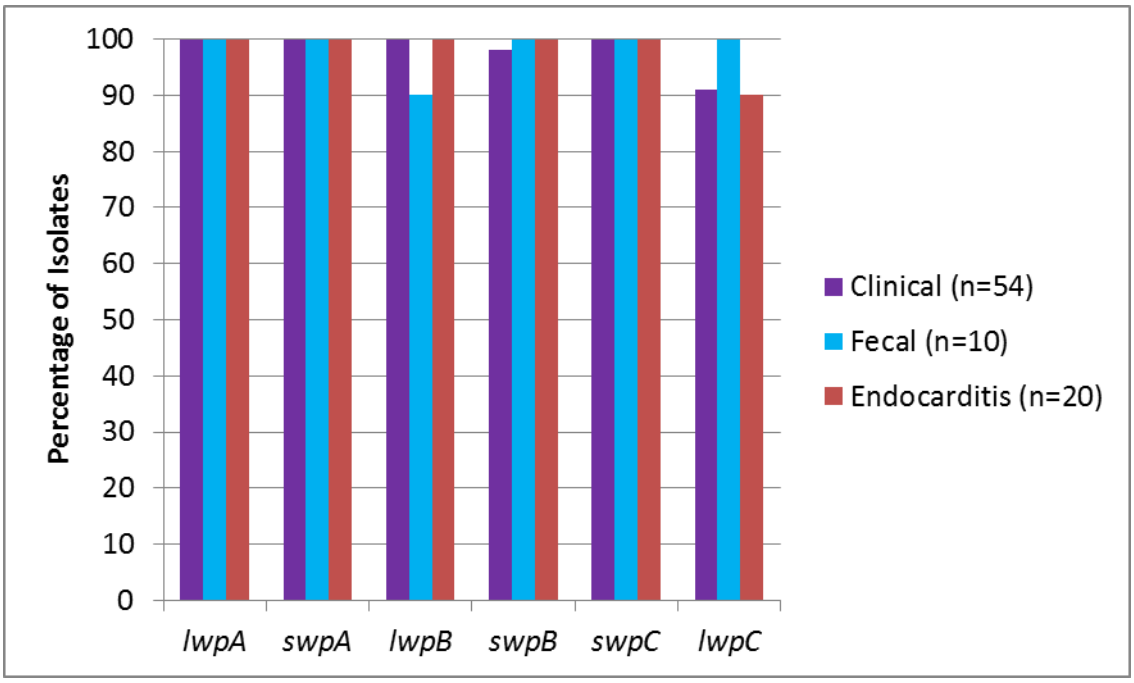


FIGURE 21. Colony hybridizations for the genes encoding WxL proteins show that these genes are ubiquitous in *E. faecium* strains and do not depend on origin. Colony hybridizations were performed with probes for each individual WxL gene using 54 clinical, 10 fecal, and 20 endocarditis isolates from our collection. The graph shows the percentage of isolates in each group that were positive for each gene.

[4.3.3] The predicted structure of the WxL proteins. The protein sequences used to make recombinant proteins of TX82 were analyzed through different protein servers; the expected molecular weights, theoretical PIs, and summary of number of aromatic residues can be found in Table 27. The recombinant proteins made from TX82 Locus A and Locus C were further analyzed for secondary structure. Using Phyre2, neither SwpA, LwpA, or SwpC could be modeled to a template with any reasonable confidence. It was found that all three were highly disordered (using DisEMBL) (data not shown), which could explain the lack of a template structure. Both DufA and DufC modeled to the same template as a protein-glutamine glutamyltransferase, specifically the three-dimensional model of the human transglutamase 32 enzyme with the highest confidence level (98.8%) using 52% and 59% of the protein sequence for DufC and DufA, respectively. LwpA modeled to a Concavalin A-like lectin with 100% confidence for 30% of the protein. This is similar to what was found in *L. plantarum* which described concavalin-A like lectin domains in the large WxL proteins (Siezen et al., 2006).

The WxL Locus A recombinant proteins purified using the sequences from TX82 were further analyzed for structure experimentally. The UV absorbance spectra from 250nm to 320nm were taken for SwpA, DufA, and LwpA (Figure 22). LwpA had the UV spectra of a pure protein (a peak at 280nm), whereas DufA and SwpA had high absorbance readings between 250nm and 260nm (although they had another peak at 280nm). High absorbance at the 260nm wavelength is normally indicative of a protein binding to DNA or contamination with DNA; however, these proteins were treated with DNase. Another characteristic that may cause the abnormal UV spectra could be the very high percentage of phenylalanine in these proteins or the binding of ions.

TABLE 27. Summary of Locus A recombinant protein characteristics.

Protein	MW	pI	Tryptophan	Tyrosine	Phenylalanine
DufA	39178	5.18	1	6	13
LwpA	74560	6.13	9	29	23
SwpA	21593	4.69	3	3	6

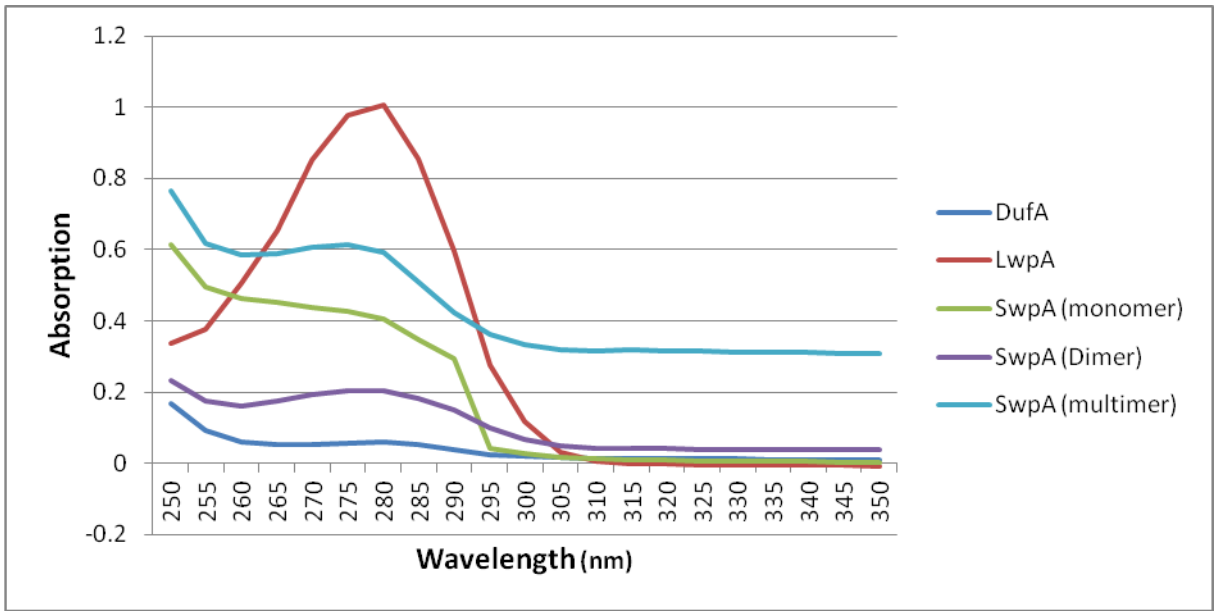


FIGURE 22. The UV absorbance spectra of purified TX82 WxL Locus A recombinant proteins. The UV absorption readings for each recombinant protein was taken at 5nm intervals from 250nm to 350nm on a spectrophotometer.

Circular dichromism was also performed for SwpA, DufA, and LwpA; the spectra and secondary structure content can be found in Figure 23 and Table 28, respectively. All three proteins contain 53-55% β-structures (sheets and turns) and 34-38% unordered structure, but very little α-helical structure (5-12%). Considering that β-sheets are aggregation drivers and proteins that bind ECM typically have disordered protein structure (personal communication with Xiaowen Liang), we can speculate a pathogenesis-related function for these proteins. Interestingly, through CD and steady-state intrinsic tryptophan fluorescence, one can see the difference in conformation/folding between the monomer and the dimer form of SwpA (Figure 23 and Figure 24). This probably explains the functional difference between the two moieties discussed later (the monomer being inactive whereas the dimer is active). Fluorescence also detected different folding of SwpA, DufA, and LwpA under different salt conditions (with speculation that these proteins might be involved in salt stress discussed later) (Figure 24).

TABLE 28. Secondary-structure content of WxL proteins determined by CD.

Protein	α-helix (%)	β-structure (%) (sheet & turn)	Unordered (%)	Total
LwpA	5.4	55.3	38.4	99.4
DufA	12.3	53.1	34.6	99.9
SwpA(monomer)	6.6	53.1	37.7	97.4
SwpA(dimer)	8.5	54.1	37.5	100

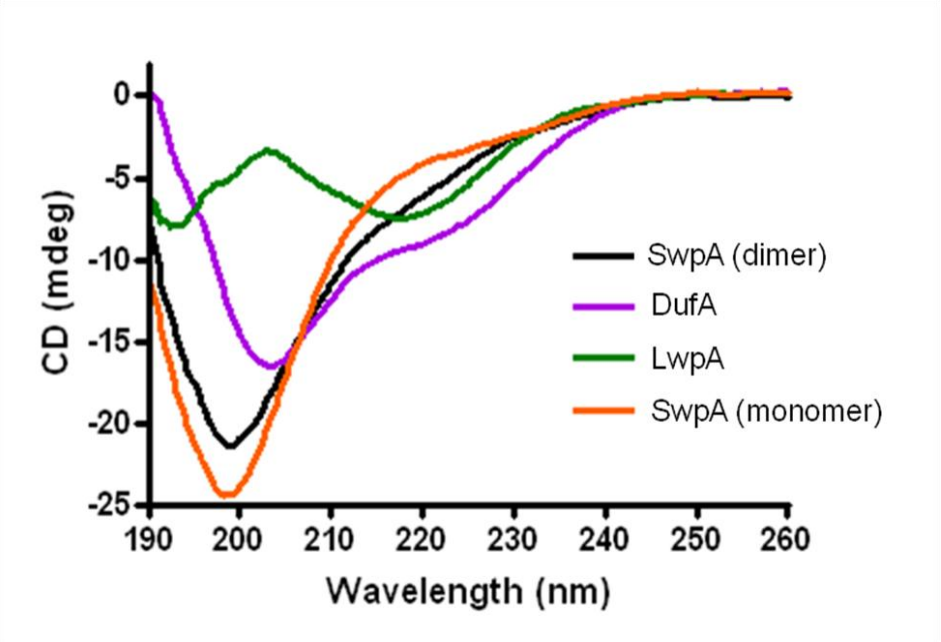


FIGURE 23. Circular dichromism spectra of purified TX82 WxL Locus A recombinant proteins. CD measurements in the far UV region (190-260 nm) were performed at ambient temperature on a Jasco J-720 spectropolarimeter with a 0.5 mm cell. Ten scans were collected and averaged at a scan speed of 100 nm/min, with a time constant of 2 s and a band-width of 1 nm.

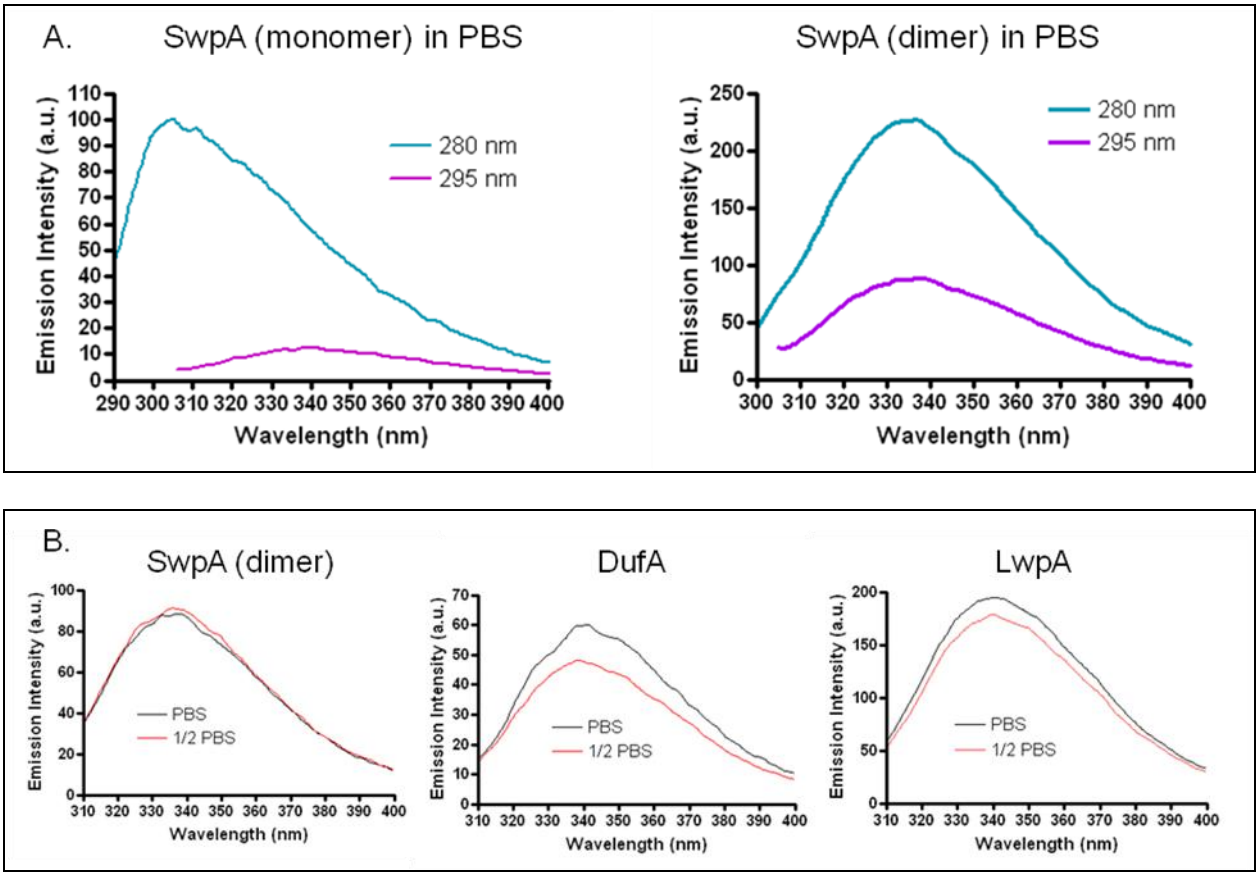


FIGURE 24. Steady-state intrinsic tryptophan fluorescence spectra show differences between the monomer and dimer forms of SwpA as well as differences under different salt concentrations for SwpA (dimer), DufA, and LwpA. Panel A depicts the different folding/conformational changes as can be measured by emission intensity in steady-state intrinsic tryptophan fluorescence between the monomer and dimer form of SwpA at the same condition. Panel B depicts the different folding/conformational changes measured by emission intensity in steady-state intrinsic tryptophan fluorescence of each individual protein under different salt concentrations (PBS and ½ PBS) at 295nm.

[4.3.4] The WxL proteins are expressed on the surface of bacterial cells, although they are differentially exposed depending on the clade of the strain. Mutanolysin was used to treat cells to obtain cell wall extracts. The cell wall extracts of TX82 cells were then run on SDS-PAGE gels and probed with specific antibodies to the WxL proteins. SwpA, LwpA, SwpC, and LwpC were all found in the cell wall fraction (Figure 25).

To confirm this, immunogold electron microscopy was performed on TX82 cells. SwpC and LwpC were expressed on the surface of TX82 cells as can be seen in the electron micrographs pictured in Figure 26. Cells labeled using the pre-immune sera gave no signal. Six HA and six CA isolates were then tested by whole-cell ELISA to determine if all isolates expressed WxL proteins at a similar level. Although all strains expressed the WxL proteins from Locus A and C on the surface of the cell, HA strains gave higher optical densities (approximately two fold higher) than CA strains for SwpA, LwpA, and LwpC (Figure 27). These differences were statistically significant. Although the results for SwpC were not statistically significant, the trend was similar to the other WxL proteins.

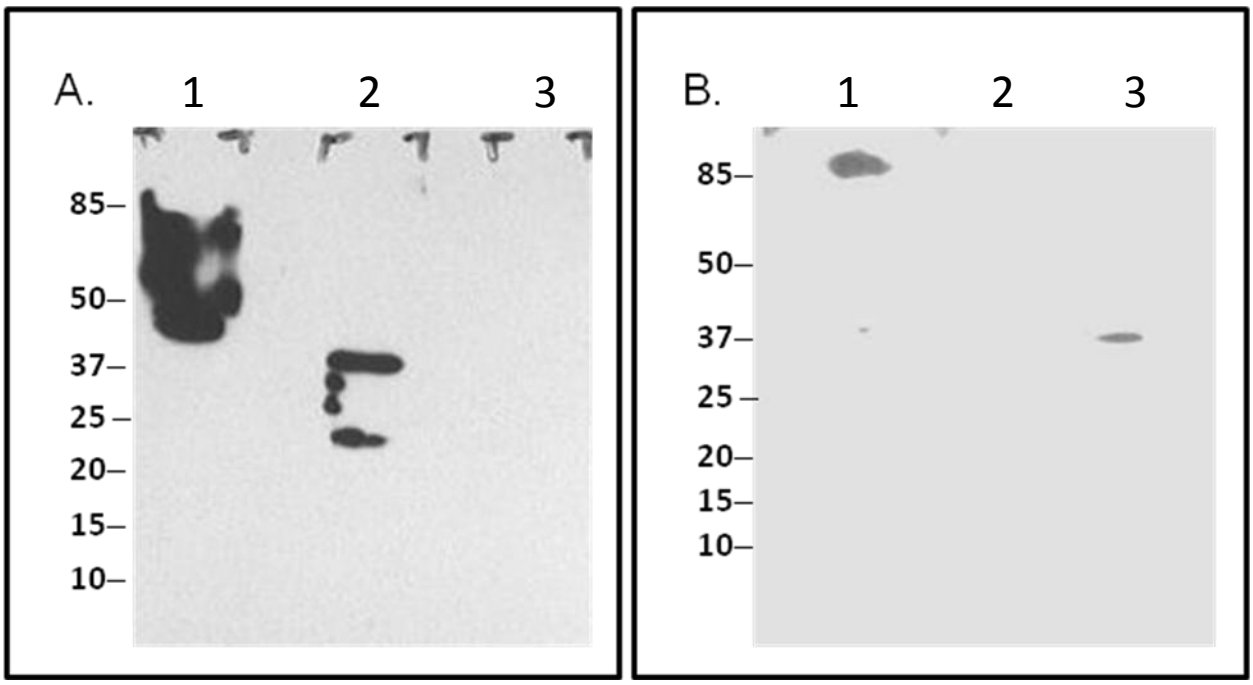


FIGURE 25. Western blots for WxL proteins of Locus A and C of mutanolysin cell wall extracts from TX82 cells. Western blots were performed and probed for SwpA, LwpA, SwpC, and LwpC in mutanolysin cell wall extracts of TX82 cells. Panel A depicts the western blot analysis for Locus A. Lanes: 1, anti-LwpA antibody; 2, anti-SwpA antibody; 3, pre-immune sera control. Panel B depicts the western blot analysis for Locus C. Lanes: 1, anti-LwpC antibody; 2, preimmune sera control; 3, anti-SwpC antibody.

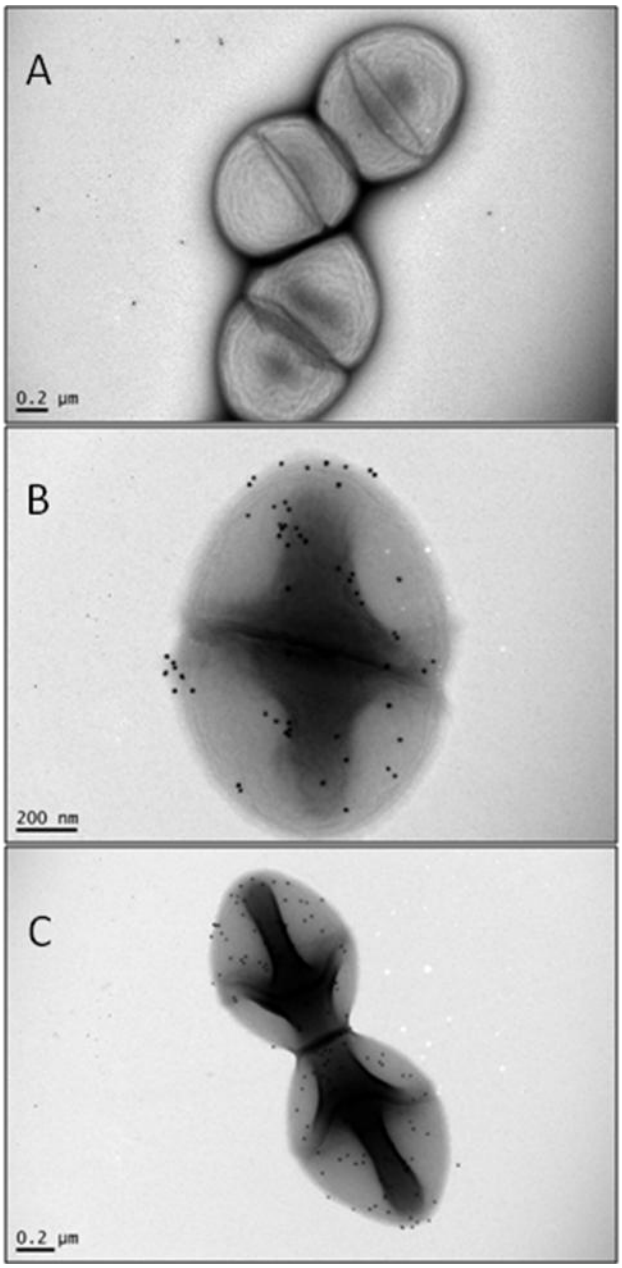


FIGURE 26. Immunogold electron microscopy shows the WxL Locus C proteins on the surface of TX82 cells. Immunogold electron microscopy was performed on TX82 cells using specific antibodies against SwpC and LwpC followed by 18nm gold conjugated anti-rat antibody. Panel A is the pre-immune sera control at a 1:50 dilution. Panel B used anti-SwpC antibody at a 1:50 dilution. Panel C used anti-LwpC at a 1:50 dilution.

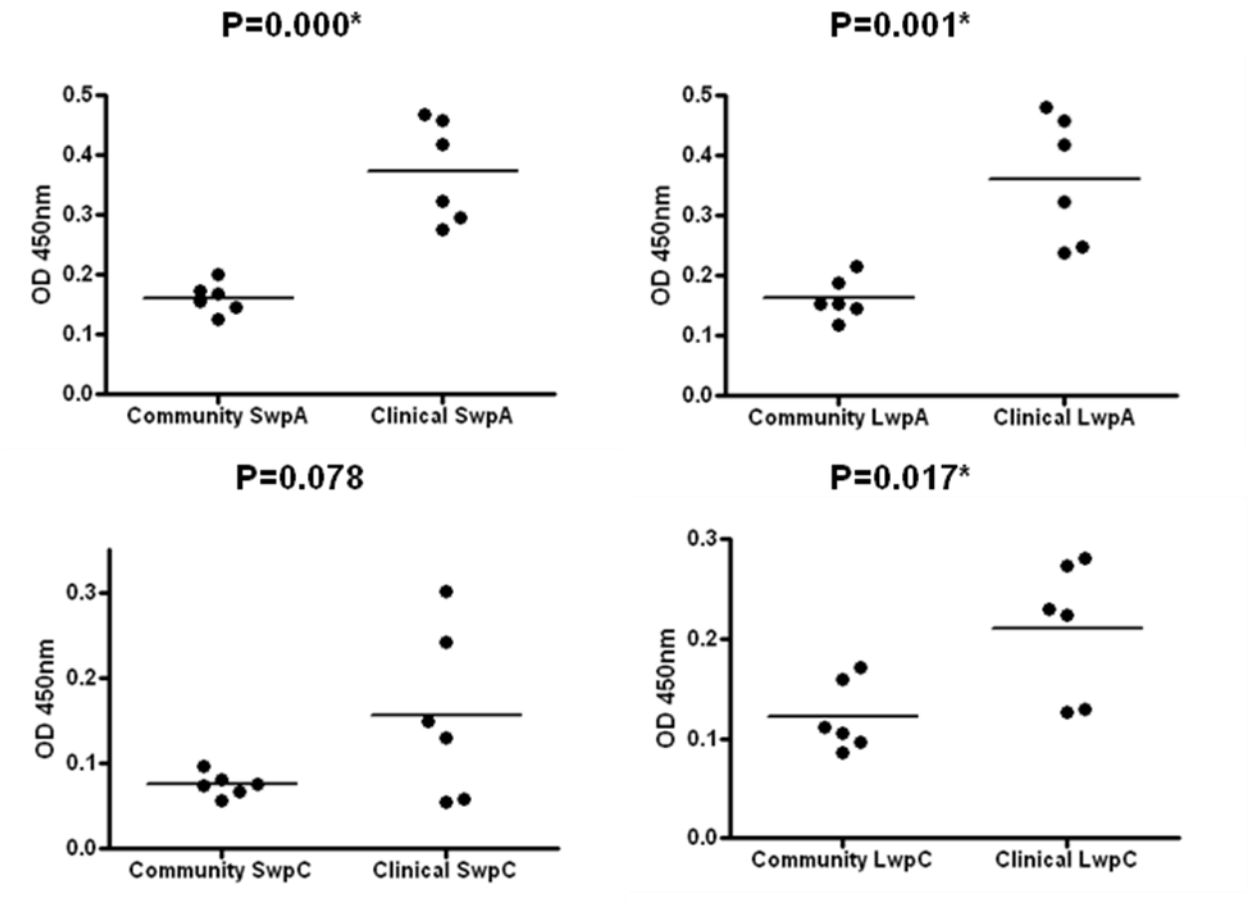


FIGURE 27. The differences in surface display/expression of WxL proteins between community clade isolates and hospital clade isolates. Whole-cell ELISA of native cells using 6 CA and 6 HA *E. faecium* strains was performed and surface display of WxL proteins detected using specific antibodies to SwpA, LwpA, SwpC, or LwpC. Standard deviation was negligible so error bars cannot be seen. Unpaired t-test was used to determine the statistical significance between the community and clinical groups for each WxL protein.

[4.3.5] Recombinant proteins from the WxL Locus A bind a variety of extracellular matrix proteins. ELISA-type ligand binding assays were performed to determine if the recombinant proteins from the WxL Locus A could bind immobilized extracellular matrix proteins (Figure 28). All three recombinant proteins SwpA, DufA, and LwpA bound a number of ECMs. SwpA bound fibronectin and fibrinogen with the highest OD readings. DufA bound collagen I, collagen IV, and laminin with the highest OD readings. LwpA bound collagen V, fibronectin, and fibrinogen with the highest OD readings.

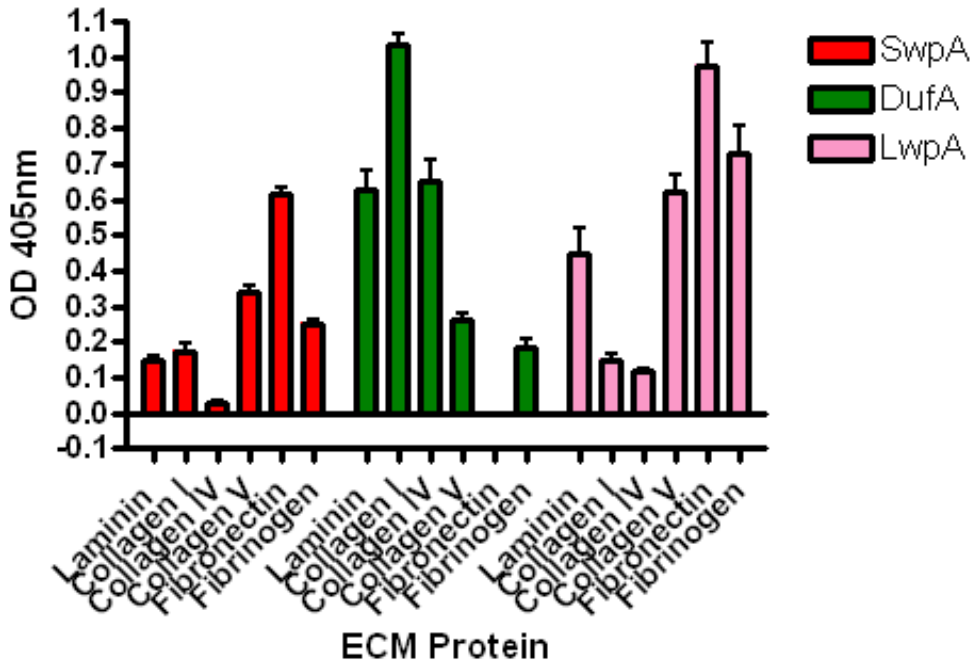


FIGURE 28. Binding of extracelluar matrix proteins by recombinant proteins from WxL Locus A. The binding of recombinant proteins from WxL Locus A (20µM) to different ECM proteins (1µg) immobilized to an ELISA plate detected with AP-conjugated anti-His antibodies. The background values of BSA bound to each ECM protein were subtracted.

Surface plasma resonance (SPR) was then performed using Biacore to screen for binding response in different ECMs, as well as to see which forms of SwpA were active (Figure 29 and Table 29). SwpA showed binding responses for collagen I, collagen IV, laminin, fibrinogen, and fibronectin. The dimer form of SwpA was the most active, showing a greater binding response in all instances than the multimer form, whereas the monomer form showed no binding to any of the ECMs. DufA also showed a binding response to collagen I, collagen IV, laminin, fibrinogen, and fibronectin. LwpA only showed binding responses for collagen IV and fibrinogen.

Concentration-dependent Biacore analysis was then used for the two ECMs, collagen I and fibronectin (also to the C-terminal Hep2-40K chain of fibronectin), to determine the apparent equilibrium disassociation constants (K_D). The binding affinities (equilibrium disassociation constants) to collagen I and fibronectin for the three recombinant proteins from Locus A are summarized in Table 29 and the response curves can be seen in Figures 30, 31, and 32. Soluble SwpA (dimer) bound both immobilized collagen I and fibronectin, as it did in the ligand binding assays. However, SwpA bound collagen I with a higher affinity than fibronectin (both when soluble and immobilized), whereas in the binding assays, fibronectin gave the highest OD reading. In addition, soluble SwpA bound full-length fibronectin with a greater affinity than the Hep2-40K chain, but did not bind the N-terminal NTD-70K chain. This differed from the result when SwpA is immobilized, where the binding affinity was greater for the Hep2-40K than full-length fibronectin.

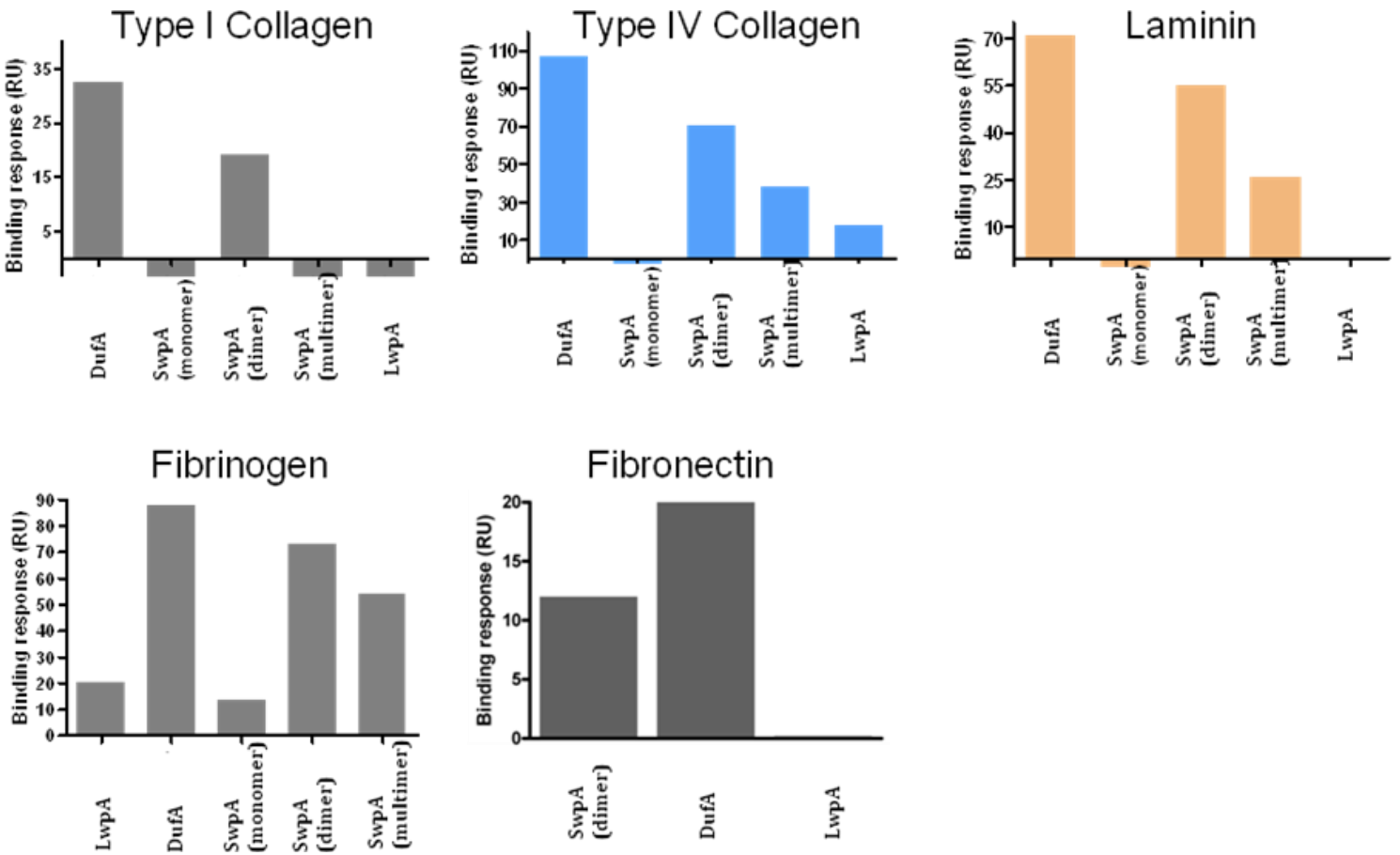


FIGURE 29. An initial screen using Surface Plasmon Resonance shows binding responses for Locus A recombinant proteins to different immobilized ECMs. Biacore was used to test for any binding response (not concentration dependent) for the Locus A

recombinant proteins to immobilized type I collagen, type IV collagen, laminin, fibrinogen, and fibronectin. For all assays, the concentration of DufA was 55µg/ml, SwpA monomer was 164µg/ml, SwpA dimer was 35µg/ml, SwpA multimer was 64µg/ml, and LwpA was 75µg/ml as determined by BCA assay. For fibronectin, only SwpA (dimer), DufA, and LwpA were used.

TABLE 29. Surface Plasmon Resonance analysis shows recombinant proteins from Locus A bind host ECM proteins.

Soluble	DufA	SwpA	LwpA	Fibronectin	Hep2-40K[c]	Collagen I
Immobilized						
Fibronectin	5	200	No			
Hep2-40K	75	600	No			
Collagen I	630	76	No			
Collagen IV	Yes	Yes	Yes			
Laminin	Yes	Yes	No			
Fibrinogen	Yes	Yes	Yes			
DufA	26	No		1400	320	5
SwpA	20	No		1900	240	5
LwpA	13	No		360	Yes	No

[a]The numbers are apparent K_D values in nM concentrations.

[b]Yes/No answers denote whether or not there was a binding response, but the KD values have yet to be determined.

[c]Hep2-40K is a specific portion of the C-terminal region of fibronectin.

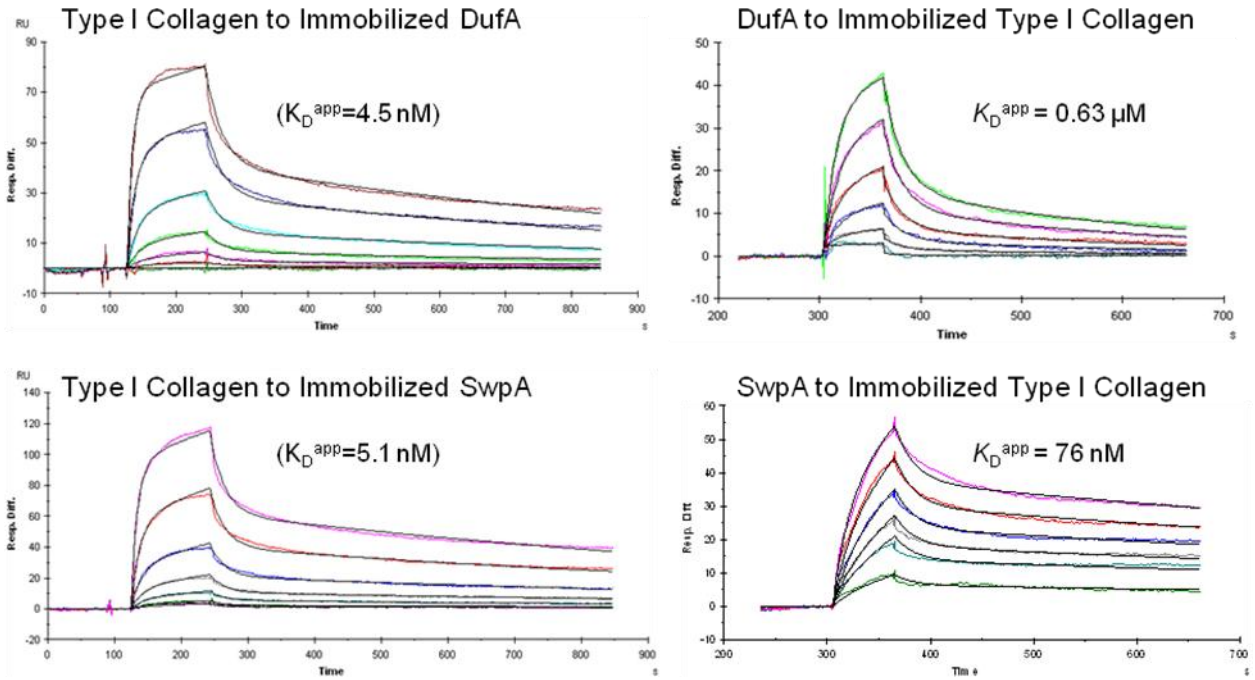


FIGURE 30. Surface Plasmon Resonance shows interactions between type I collagen, DufA, and SwpA. Biacore sensorgrams were generated (shown in black) with the lower curve corresponding to lowest concentration of soluble protein injected. Kinetic analysis (the fitted curves are shown in colored lines) was performed to obtain the rate constants, specifically the K_D^{app} values. Concentrations for soluble human type I collagen binding to immobilized DufA and SwpA ranged from 1-64nM. Concentrations for soluble DufA binding to immobilized human type I collagen ranged from 0.05-1.6µM. Concentrations for soluble SwpA binding to immobilized human type I collagen ranged from 25-800nM.

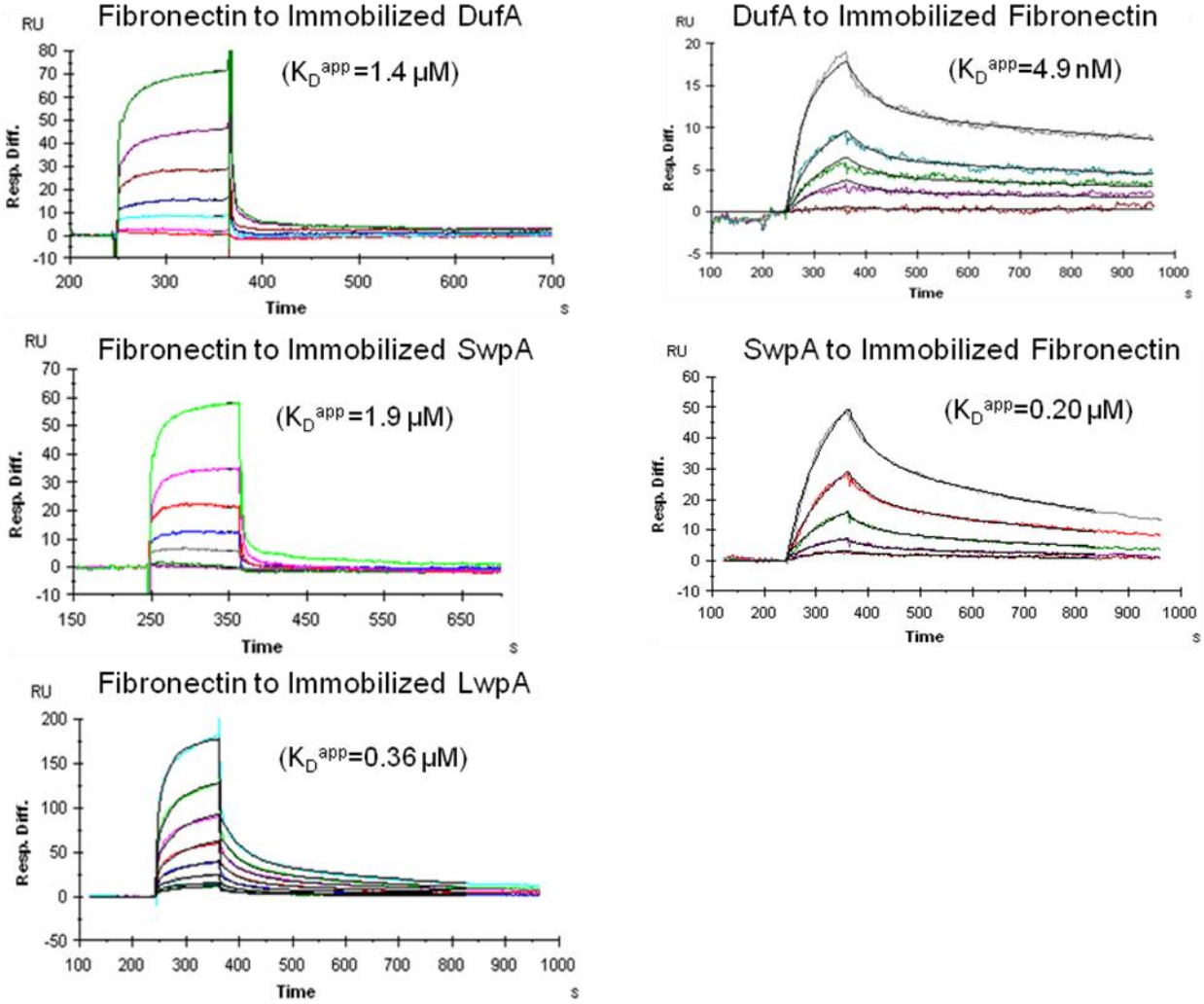


FIGURE 31. Surface Plasmon Resonance shows interactions between fibronectin, DufA, SwpA, and LwpA. Biacore sensorgrams were generated (shown in black) with the lower curve corresponding to lowest concentration of soluble protein injected.

Kinetic analysis (the fitted curves are shown in colored lines) was performed to obtain the rate constants, specifically the K_D^{app} values. Concentrations for soluble fibronectin binding to immobilized DufA, SwpA, and LwpA ranged from 0.01-1.28 µM. Concentrations for soluble DufA binding to immobilized fibrinogen ranged from 5-80nM. Concentrations for soluble SwpA binding to immobilized fibrinogen ranged from 0.025-0.4 µM.

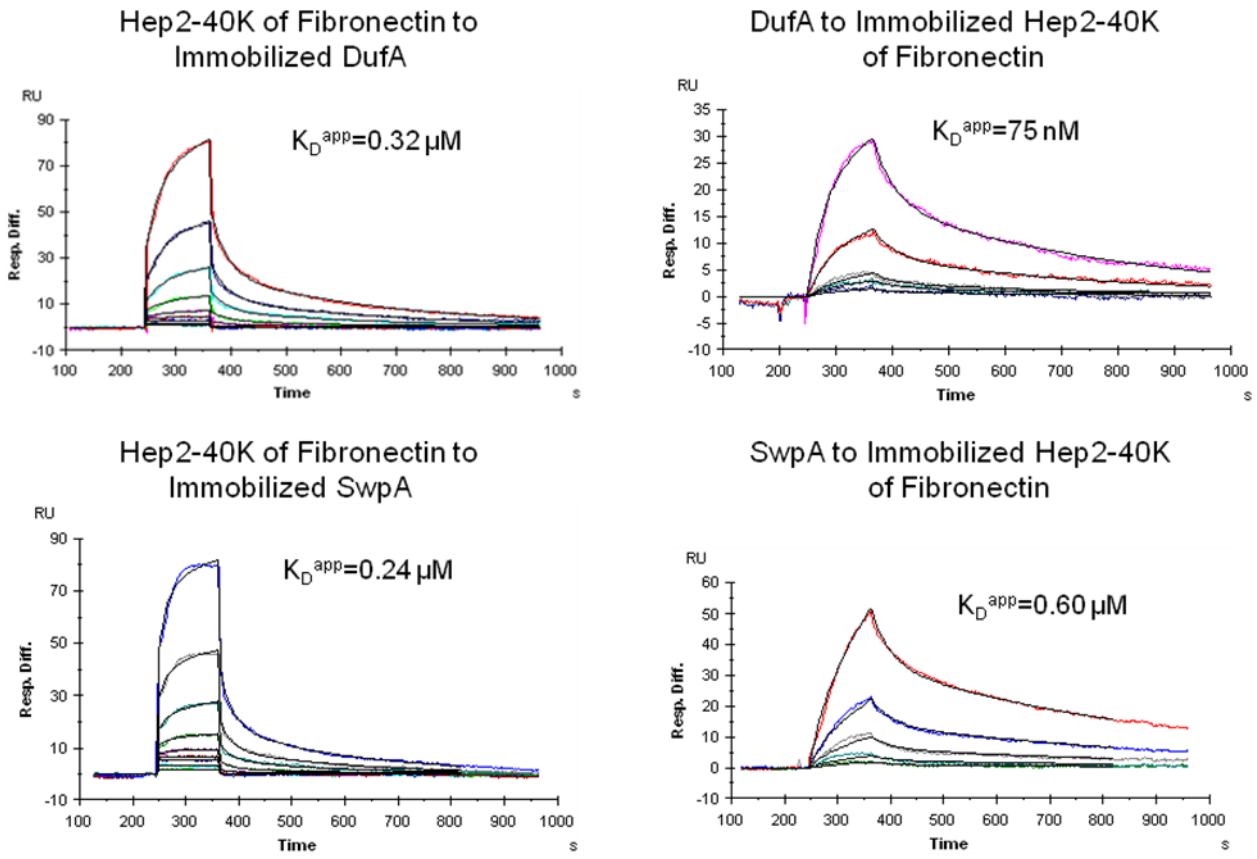


FIGURE 32. Surface Plasmon Resonance shows specific interactions between of DufA and SwpA with the Hep2-40K chain of fibronectin. Biacore sensorgrams were generated (shown in black) with the lower curve corresponding to lowest concentration of soluble protein injected. Kinetic analysis (the fitted curves are shown in colored lines) was performed to obtain the rate constants, specifically the K_D^{app} values. Concentrations for soluble Hep2-40K chain of fibronectin binding to immobilized DufA and SwpA

ranged from 1-256nM. Concentrations for soluble DufA binding to immobilized fibrinogen ranged from 5-160nM. Concentrations for soluble SwpA binding to immobilized fibrinogen ranged from 0.025-0.4 µM.

Soluble DufA also bound both immobilized collagen I and fibronectin (Figures 29-32). Interestingly, DufA bound fibronectin with a higher affinity than collagen using Biacore, but bound collagen with high OD and didn't bind fibronectin at all in the ELISA ligand-binding assays. On the other hand, when DufA was immobilized, it bound collagen I at a greater affinity than fibronectin. Similar to SwpA, soluble DufA bound the full-length fibronectin with a greater affinity than the Hep2-40K chain, and did not bind the N-terminal NTD-70K chain. Also similar to SwpA, immobilized DufA had a greater binding affinity for the Hep2-40K than full-length fibronectin.

Interestingly, you can see that the binding affinities are a result of different kinetics depending on the ECM protein used or which protein was immobilized or soluble. Soluble type I collagen binding to both immobilized DufA and SwpA had similar kinetics and binding affinites. However, when collagen was immobilized, the binding affinity to SwpA was stronger due to the increase in the stability of the reaction (slow dissociation), despite the association/recognition being slower (Figure 30).

Fibronectin binding to immobilized Locus A proteins resulted in similar binding affinities for DufA and SwpA where the association constant (K_a) was relatively high; however, the binding was very unstable as can be seen by the very quick dissociation. There was a slightly better affinity for the binding of fibrinogen to LwpA compared to DufA and SwpA, which seems to be a result of a slower dissociation, despite slower recognition/ association. The binding of recombinant DufA and SwpA resulted in a much better binding affinity as a result of slower dissociation than when fibronectin was soluble (Figure 31). Similar kinetics were observed when comparing the soluble Hep2-40K portion of fibronectin to immobilized Hep2-40K portion of fibronectin, as was seen with the full-length fibronectin (Figure 32).

[4.3.6] DufA shows self-association and association with both SwpA and LwpA. Because it was hypothesized that the proteins from the same operon might make a cell surface complex, Biacore was used to detect binding between the recombinant proteins from Locus A. As seen in Table 30 and Figure 33, soluble DufA could bind immobilized SwpA, LwpA, and itself. However, if DufA was immobilized, it could only bind itself. SwpA and LwpA could only bind DufA if DufA was immobilized, but not when DufA was soluble, and they could not bind each other. Also, DufA binding response seemed to be affected by salt concentration, as it only produced a binding response in ½ PBS buffer.

When comparing the kinetics of the interactions (although the binding affinities seem to be relatively similar for all three protein-protein interactions when DufA is soluble), it seems that the interaction with LwpA is the most stable as it dissociated the slowest, whereas SwpA dissociates the quickest. SwpA, however, does seem to have the quickest association/ recognition to DufA.

TABLE 30. Surface Plasmon Resonance analysis shows recombinant proteins from Locus A bind each other.

Soluble	**DufA**	**SwpA**	**LwpA**
Immobilized			
DufA	26	No	No
SwpA	20	No	Yes
LwpA	13	No	Yes

[a]The numbers are apparent K_D values in nM concentrations.

[b]Yes/No answers denote whether or not there was a binding response, but the KD values have yet to be determined.

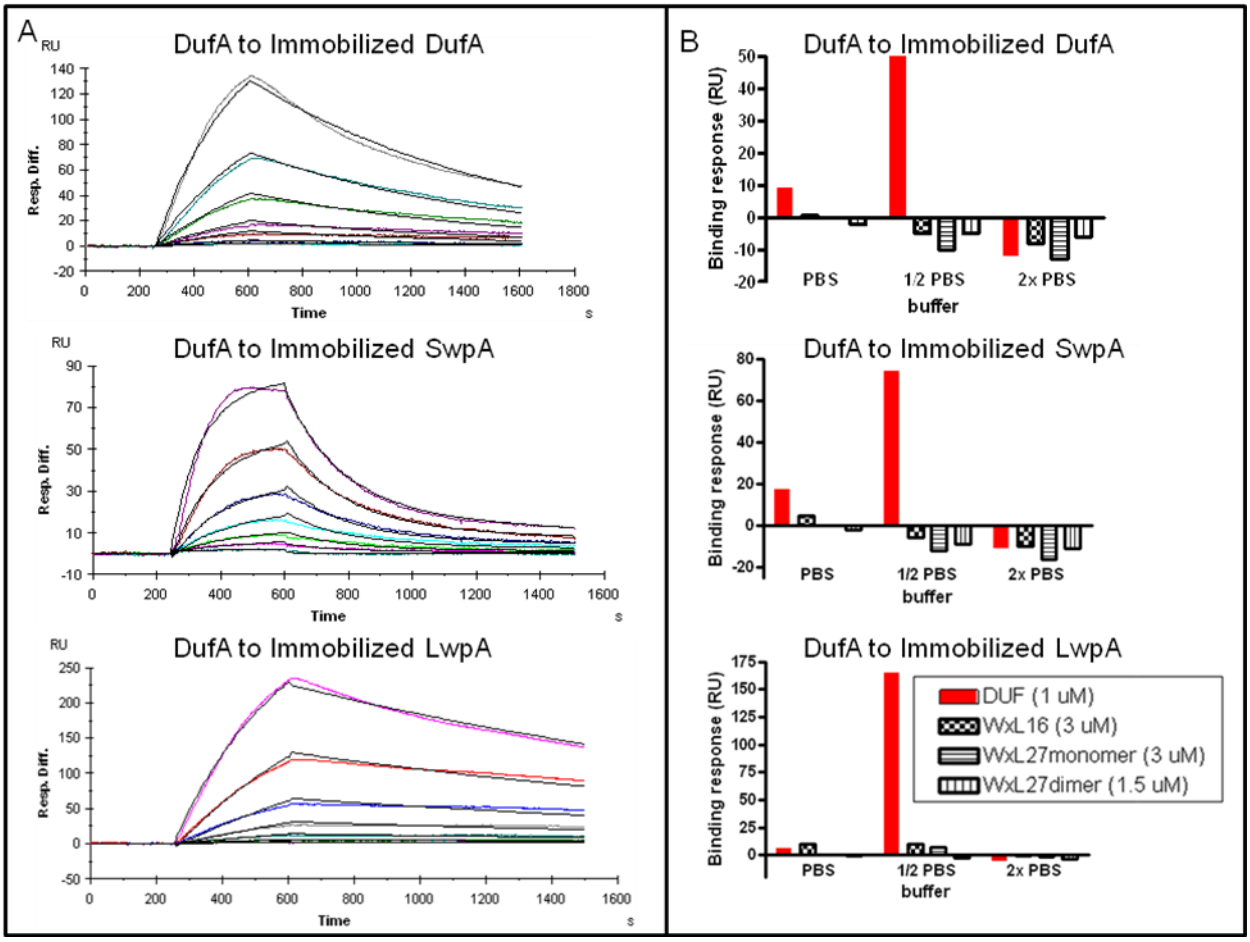


FIGURE 33. Surface Plasmon Resonance shows specific interactions between DufA with itself, SwpA and LwpA that is dependent on salt concentration. Panel A. Biacore sensorgrams were generated (shown in black) with the lower curve corresponding to lowest concentration of soluble protein injected. Kinetic analysis (the fitted curves are shown in colored lines) was performed to obtain the rate constants, specifically the K_D^{app} values. Concentrations for soluble DufA binding to immobilized DufA, SwpA, and LwpA ranged from 1-64nM. Panel B. SPR analysis using Biacore to determine if salt concentration affects the binding response of DufA to itself, SwpA, and LwpA.

[4.3.7] WxL Locus A and WxL Locus C mutants are attenuated in type I collagen binding, but not type V collagen binding or biofilm formation. Nonpolar, markerless deletions were made of each putative WxL operon in TX82. Since recombinant proteins from Locus A were shown to bind collagen I, the WxL Locus A and WxL Locus C mutants were tested for their ability to bind type I collagen. As seen in Figure 34, the WxL Locus A and WxL Locus C mutants were attenuated in binding to type I collagen, but not type V collagen compared to the parent strain TX82. These attenuations were statistically significant and not due to a growth defect (data not shown).

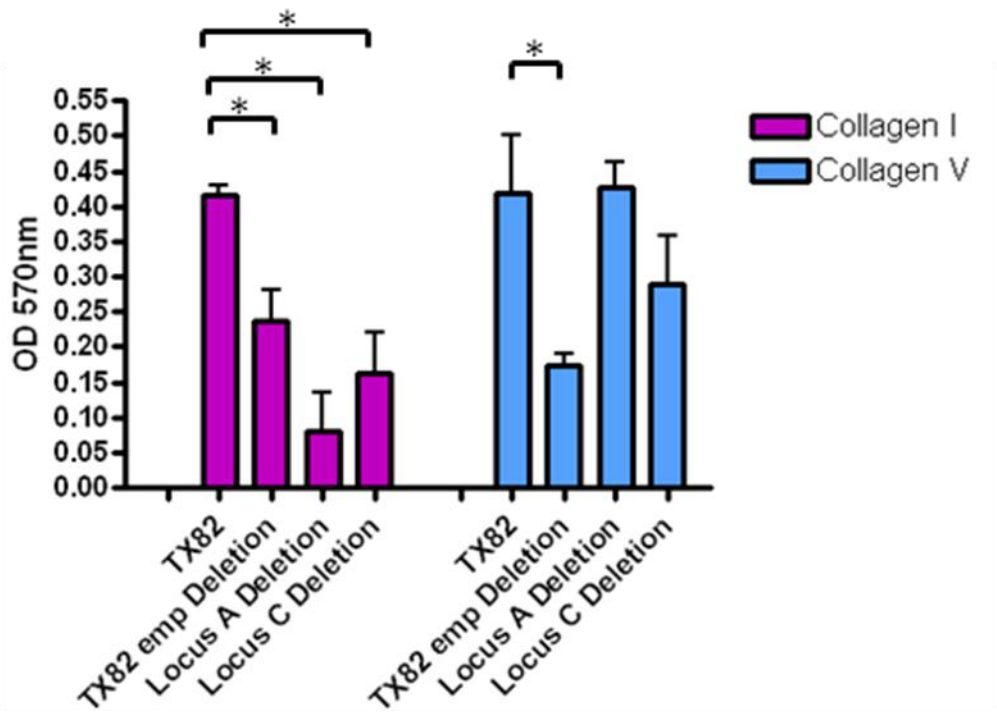


FIGURE 34. The WxL Locus A and WxL Locus C deletion mutants are attenuated in binding to type I collagen. Whole-cell ELISA analyzing the binding of WxL Locus mutants versus the parent strain (TX82) to collagens I and V (1µg). Cells were detected using crystal violet. TX82 *empABC* deletion is a mutant of a pilus operon which has known attenuation in ECM binding (unpublished data). Asterisks indicate a statistically significant difference. Type

V collagen was used as a negative control, as recombinant WxL proteins were not shown to bind this ECM.

We next tested the mutants for biofilm formation. However, neither the Locus A or Locus C mutant were attenuated in biofilm formation compared to the parental strain, as can be seen in Figure 35.

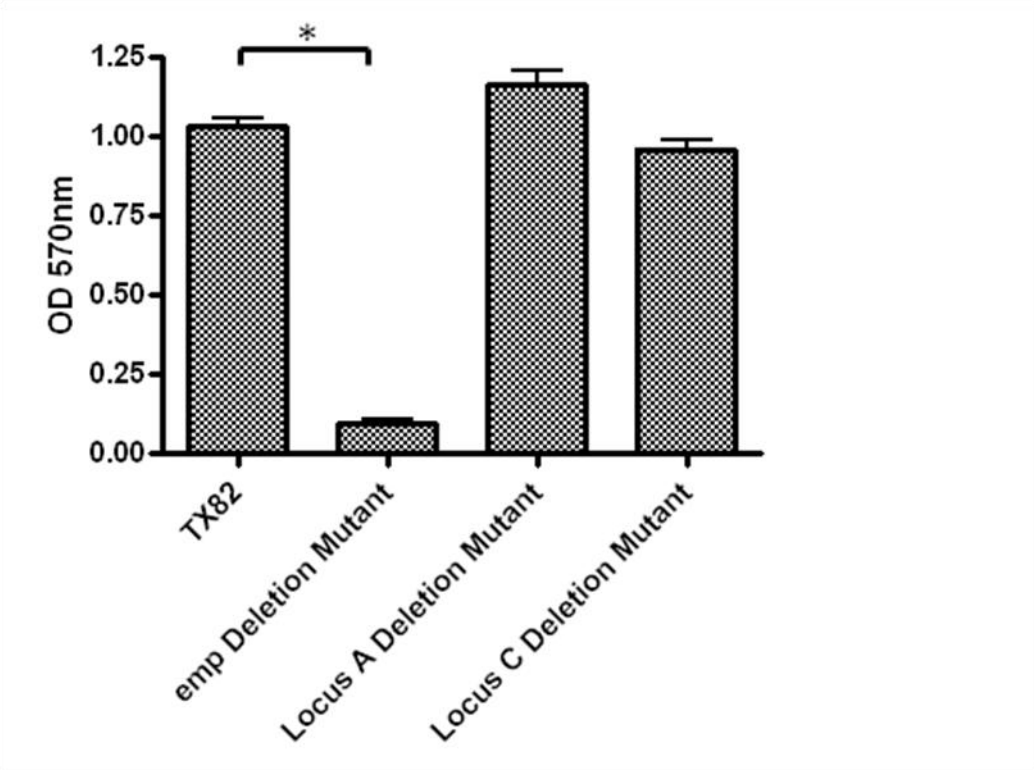


FIGURE 35. The WxL Locus A and C mutants are not attenuated in biofilm formation. Biofilm formation of TX82 and WxL Locus mutants was assayed using biofilm cells fixed with Bouin's fixative and stained with 1% crystal violet solubilized using ethanol/acetone (80:20) and absorbance was read at 570nm. The *empABC* deletion mutant was used as a control.

[4.3.8] WxL Loci could be involved in bile and salt stress response. After analyzing the transcriptomic analysis of lactobacilli, it was found that their WxL genes are regulated in response to bile, salt, and lactate stress (Siezen et al., 2006). Of interest, different salt concentrations changed the conformation of the recombinant proteins from WxL Locus A change and DufA's interaction with immobilized SwpA and LwpA was also affected by salt concentration (Figure 24 and 33). Additionally, the Locus B mutant was tested in the bile salt assay and it was found that the Locus B mutant survives better in bile salt stress conditions than the parental strain (Figure 36).

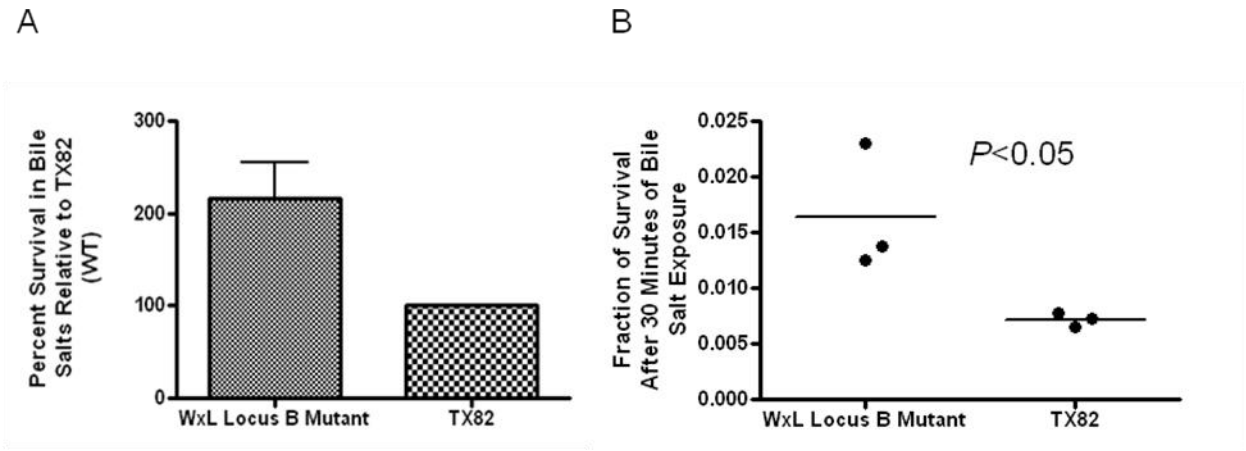


FIGURE 36. The WxL Locus B mutant survives better that the wild-type after bile salt exposure. *E. faecium* cells were exposed to BHI with 0.3% bile salt concentration for 30 minutes. CFUs were determined at t=0 minutes and t=30 minutes for the WxL Locus B mutant and TX82 (WT) in three separate experiments. Panel A shows the percent survival after 30 minutes of exposure to bile of the WxL Locus B mutant relative to the wild-type. Panel B shows the fraction of cells that survived after 30 minutes of bile salt exposure compared to t=0 for both the WxL Locus B mutant and TX82. An unpaired t-test determined the difference in mean survival between the two to be statistically significant.

[4.3.9] WxL proteins are antigenic, and may have use in protection against *E. faecium* infections. One of the aims of this study was to determine if WxL proteins are antigenic during infection. A way to test this is to determine if patients with *E. faecium* infection have antibodies to WxL proteins. Serum was previously collected from five patients with endocarditis due to *E. faecium*. Using western blot analysis, we were able to determine that endocarditis patients had antibodies to all four WxL proteins analyzed (SwpA, LwpA, SwpC, and LwpC) as can be seen in Figure 37A. This differs from healthy volunteers who either lack antibodies to any of the WxL proteins (3/7), or have antibodies to only 1 of the 4 WxL proteins. Interestingly, this was always to one of the large WxL proteins, three to LwpA and one to LwpC (Figure 37B).

We next actively immunized rats with an equal mix of LwpA and LwpC in order to determine if prophylactic immunization with the two large WxL proteins could be protective against *E. faecium* endocarditis infections. In order to confirm that one antigen did not produce significantly more antibodies than the other, serum was taken from the immunized rats to perform antibody titers (data not shown). All rats immunized had an antibody titer >1:12800 for each antigen.

In the first experiment when immunized rats were challenged with *E. faecium* endocarditis infection with TX82 (Figure 38), 7 of the 8 rats that were immunized were uninfected, compared to non-immunized rats (2), which had 10^5 - 10^6 CFU/gm of vegetation. However, 2 of the 3 sham rats (only immunized with FCA/FICA, no recombinant protein) had low levels of bacteria (between 10^1 and 10^2) with only one sham rat having CFU counts relative to the non-immunized. A second experiment was done to see if the sham effect was real or a technical error. Unfortunately, in this experiment the opposite of the desired effect occurred. Non-immunized rats again had average CFU/gm of vegetation around 10^6. However, the

immunized rats had a higher average CFU/gm than those immunized with the adjuvant alone. When the two experiments were combined, although there was a statistically significant difference between the non-immunized and immunized groups, it appears that this is due to the adjuvant instead of the recombinant proteins themselves.

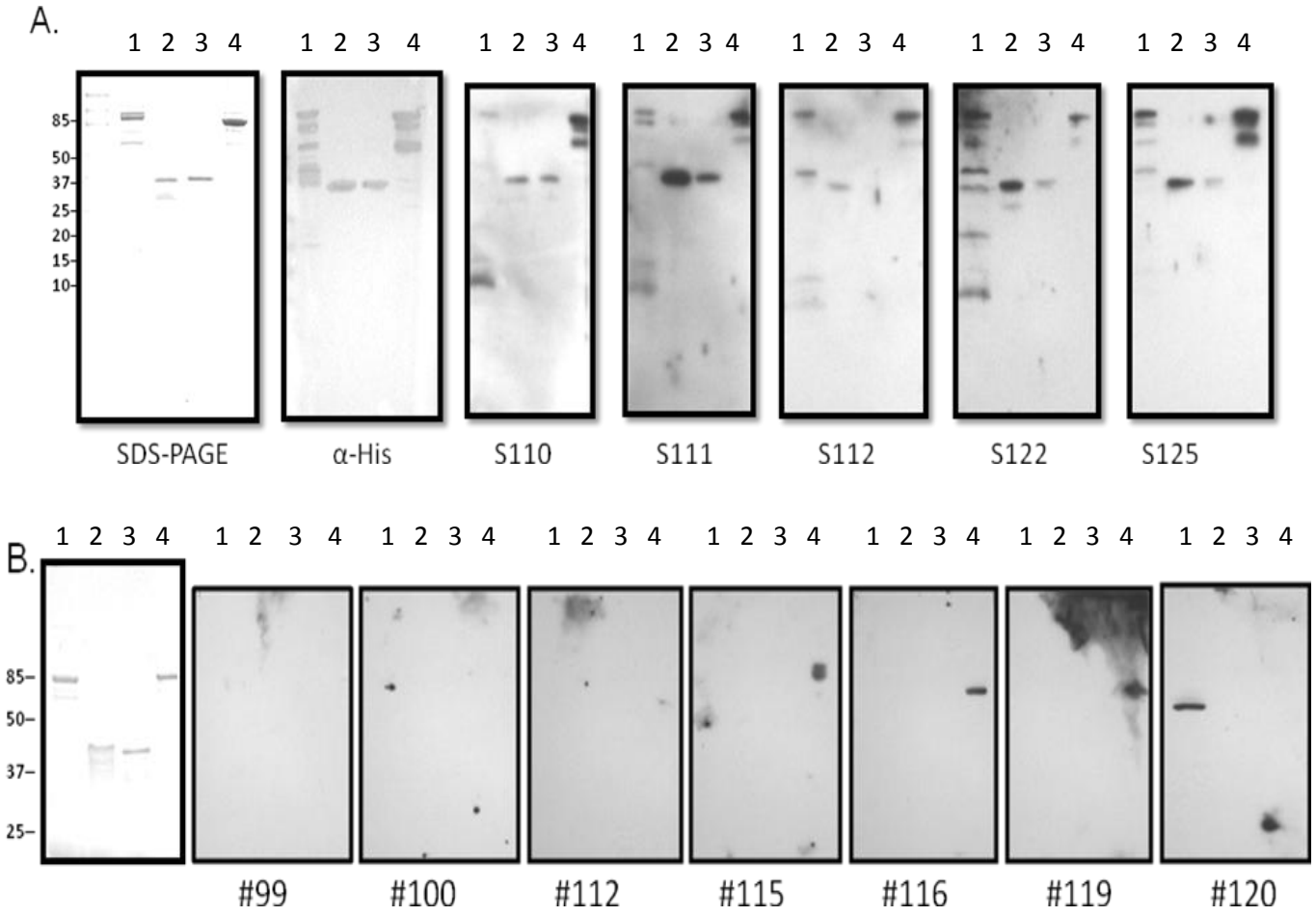

A.
1 2 3 4
85-
50-
37-
25-
20-
15-
10-
SDS-PAGE
α-His
S110
S111
S112
S122
S125
B.
85-
50-
37-
25-
#99
#100
#112
#115
#116
#119
#120

FIGURE 37. The WxL proteins of Locus A and Locus C are antigenic in patients. Recombinant proteins (Lanes: 1, LwpA; 2, SwpA; 3, SwpC; 4, LwpC) were run on 12% SDS-PAGE gels and transferred to PVDF membranes. Each membrane was incubated with either endocarditis patient serum (Panel A) or healthy volunteer serum (1:2000) (Panel B). This was followed by goat anti-human IgG conjugated to HRP and developed using Supersignal West Pico chemiluminescent substrate (Thermo Scientific). The first gel of each panel is a commassie stained gel. The numbers underneath each gel represent the patient number arbitrarily assigned.

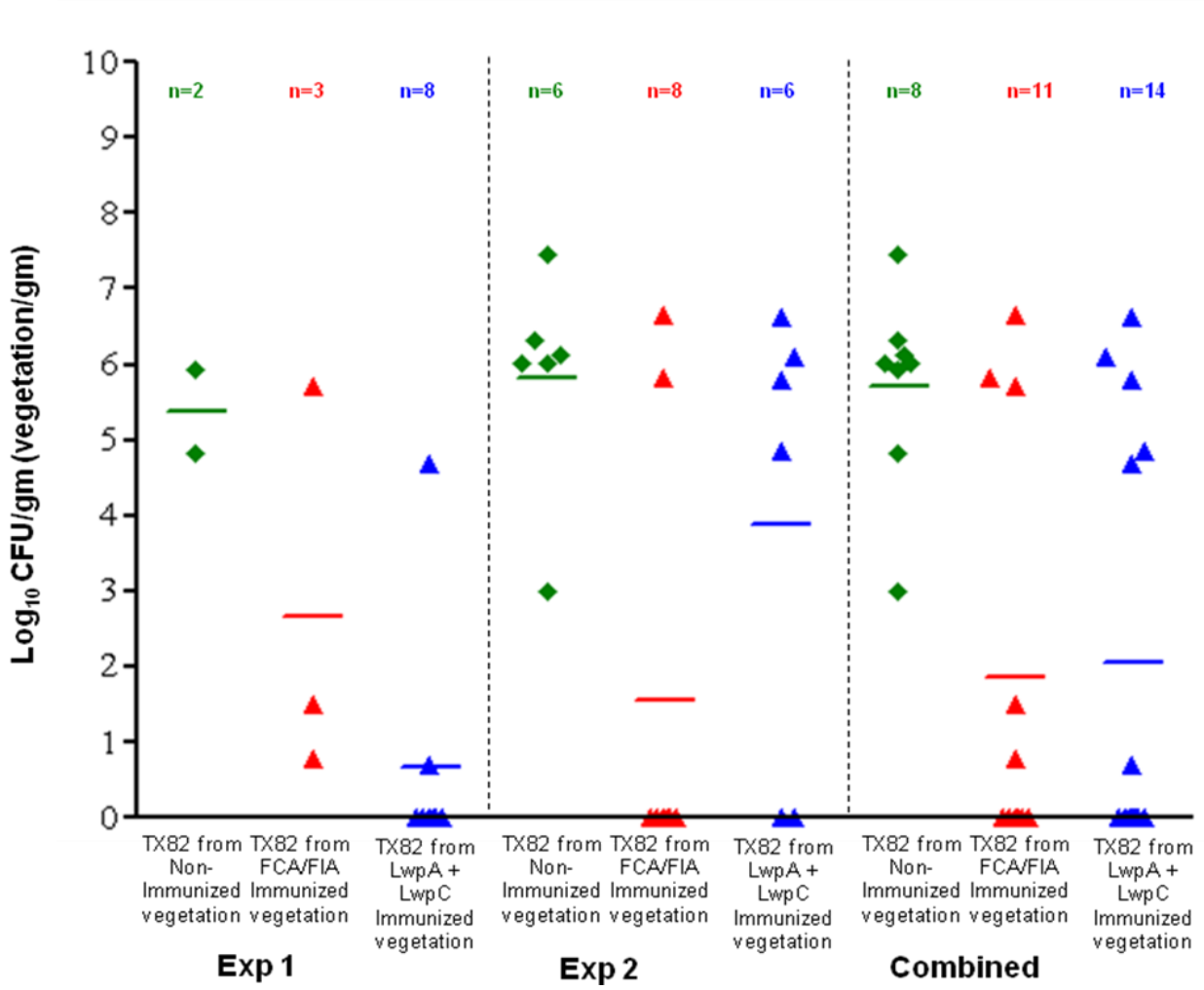


FIGURE 38. Protection against TX82 endocarditis after active immunization with LwpA and LwpC.

Rats were actively immunized with a mix of LwpA and LwpC, a sham control (FCA and FICA only), or not at all. After heart catheterization, animals were challenged with 10^9 cells of TX82 via tail vein injection. After 48 hours, vegetations were collected and CFUs determined. Differences in the bacterial $\log_{10}$ CFU (geometric mean) of the heart vegetations from the three groups were analyzed by unpaired *t*-tests. The Fisher's exact test was then used to compare the total number of infected vs. non-infected rats in the three groups.

[4.4] CHAPTER 3 DISCUSSION

Although proteins with the WxL domain have been found in 17 low-G+C content Gram positive species, with up to 27 proteins in *E. faecalis*, they have only been characterized in a few bacterial species (Brinster et al., 2007a; Brinster et al., 2007b; Schachtsiek et al., 2004; Siezen et al., 2006). Even so, these studies are limited, only describing the arrangement of the genes and the possible role of the WxL domain, with inadequate characterization of the full length proteins and little experimental data to support the proposed function of these genes (Brinster et al., 2007a; Brinster et al., 2007b; Schachtsiek et al., 2004; Siezen et al., 2006).

Brinster and colleagues reported the first characterization of an *E. faecalis* WxL protein encoded by the EF2686 locus, which they named ElrA. Inactivation of the *elrA* gene significantly reduced virulence in the mouse peritonitis model as well as reduced dissemination in the liver, spleen, and peritoneal macrophages. Furthermore, it was shown that ElrA has a role in eliciting the IL-6 response (Brinster et al., 2007b). In addition, another group identified an ElrA homolog in *Lactobacillus coryniformis*, named Cpf, and found that it mediated coaggregation with other bacterial species (Schachtsiek et al., 2004). To our knowledge, these are the only functional studies of WxL proteins excluding genomic and transcriptional analysis.

When looking at genome distribution, these clusters are also found in *Listeria*, *Lactococcus*, *Pediococcus*, and *Bacillus* species, but are interestingly missing from *Streptococcus*, *Staphylococcus*, and *Clostridium* species, suggesting these proteins may be needed for a specific niche (Siezen et al., 2006). Between the species that contain these clusters, the number of clusters varies widely from nine clusters in *L. plantarum* to one cluster in *Bacillus* species. Moreover, not all WxL proteins are contained in clusters (Siezen et al., 2006).

In *E. faecium*, sequence analysis determined that the nucleotide and amino acid identity between the alleles of different strains ranged from 86-100%. Interestingly, this differs from *E. faecalis* strains analyzed (OG1RF and V583) where the identity ranged from 21-88% (Bourgogne et al., 2008). The relative conservation of these genes in *E. faecium* may be correlated with the relative low number of these genes in this species compared to *E. faecalis* V583 which contains eight clusters and 27 WxL genes (Brinster et al., 2007a), and *L. plantarum* WCSFI with nine clusters and 22 WxL proteins (Siezen et al., 2006).

The *E. faecium* WxL genes were found to be ubiquitous throughout the genomes (Figure 21) and strains studied by genomic analysis or colony hybridization, regardless of the origin of the strain. Therefore, these are different from some of the other surface proteins studied in the previous chapters, which are enriched in the hospital clade. Although these genes segregated into the two clade pattern found in Chapter 2 (Figure 19), there was a larger difference between the CA and HA clade strains in their WxL alleles compared to other genes analyzed in the genome (Galloway-Pena et al., 2012; Galloway-Pena et al., 2011). This is possibly due to the fact that they are surface proteins and seem to be antigenic, as discussed below. Moreover, even though the genes are ubiquitous among the strains studied, there was a difference in surface expression/exposure of the WxL proteins depending on the origin of the strain (Figure 27). These differences in surface exposure could be due to the variation in the sequences we see between the CA and HA clade, resulting in altered expression or processing at the cell surface. In addition, this greater surface exposure could be reflected in a greater capacity for virulence (i.e., better able to bind host ECM).

The RT-PCR and *in silico* analysis indicate that WxL proteins exist in operons in *E. faecium* (Figure 17 and Figure 18). The presence of a small LPxTG protein within these

operons (although only Locus A in *E. faecium*) may indicate some novel assembly process at the cell surface. Indeed, the WxL proteins were found in cell wall extracts (Figure 25) and could be visualized on the surface of the cell (Figure 26). Although immunogold micrographs using double labeling with two different sized gold particles did not show colocalization of SwpC and LwpC (data not shown), we cannot rule out that they co-localize as the same secondary antibody was used for detection (anti-rat). Labeling the proteins using fluorescent markers and fluorescence microscopy could be used to better detect co-localization. Also in support of assembly of these proteins, is the binding of DufA to itself, SwpA, and LwpA as seen by SPR (Figure 33). Indeed, aggregates were seen in the waste solution after Biacore analysis (a mix of the three recombinant proteins in solution) also suggesting their association with each other (data not shown). However, the specific interactions of the proteins encoded within the same operon would need to be confirmed using methodologies such as pull down assays. Interestingly, although an LPxTG-like protein, which is covalently linked to the cell surface, is found within Locus A, it is assumed that the WxL proteins are non-covalently linked to the bacterial cell surface. The WxL proteins, at least in lactobacilli, can be dispersed from the cell surface via LiCl and reattached (Schachtsiek et al., 2004).

Looking at the predicted structures of the WxL Locus A and C proteins, it is apparent that there are a number of domains and characteristics that would lead one to speculate that these proteins may be involved in specific functions. The concavalin-A like lectin domain found in LwpC, apparently missing from LwpA, could support the speculation of involvement in carbohydrate metabolism or catabolism in general. Indeed, WxL proteins are regulated by the CcpA (catabolite control protein A) in lactobacilli (Siezen et al., 2006). Concavalin-A binds structure on sugars, glycoproteins, and glycolipids, and can bind metallic ions, which perhaps

explains the conformation, folding changes, and different binding affinities seen in different salt concentrations (Figure 24 and Figure 33). Concavalin A has also been reported to interact with immunoglobulins and agglutinate erythrocytes, also suggesting a pathological function for the WxL proteins harboring this domain.

The presence of β-structures and disordered domains also leads to speculation regarding pathogenicity function. The two highly ordered β-sheets with a linker in between seen in the predicted structures of DufA, DufC, and LwpC are very similar to the structure of the binding domains of MSCRAMMs (Hendrickx et al., 2009b). In addition, MSCRAMMS also have highly disordered regions, which leads to their ability to bind different host ECMs; SwpA, DufA, LwpA, and SwpC are predicted to also have these regions (data not shown).

Indeed, we found that the recombinant proteins SwpA, DufA, and LwpA bind a variety of extracellular matrix proteins (Figure 28 thru Figure 32). ELISA-type ligand binding assays were used as an original screen to detect binding to ECMs. When binding to host ECM was further investigated through SPR, there were some differences between the ELISA-type assay and the SPR analysis. SwpA gave binding responses using Biacore analysis for type I collagen, type IV collagen, laminin, fibrinogen and fibronectin, whereas in the ELISA-like ligand binding assay it did not bind type IV collagen. DufA bound all ECM proteins tested by SPR as well, but did not bind fibronectin in the ELISA-like assay. These discrepancies are probably due to the differences in nature of the two assays. SPR is more sensitive and can detect more transient or "quick" interactions, whereas in the ELISA-type ligand binding assay, the recombinant proteins incubate over a 1-2 hour period before binding is measured. LwpA is the opposite in that it bound all ECMs tested in the ELISA-type assay, but did not give binding responses for type I collagen, laminin, or fibronectin using SPR. This discrepancy could be due

to the differences in folding of LwpA in different solutions; perhaps it needs a particular pH or salt concentration to bind host ECM; the buffers used for the two assays had different pH and salt concentration.

In addition, differences were seen in the binding affinities depending on whether the ECM analyzed was soluble or immobilized (Table 29). For both DufA and SwpA, the binding affinity was greater when type I collagen is soluble. This is probably due to the fact that there are more binding sites when type I collagen in soluble and therefore the bound recombinant proteins can bind to multiple places in a single collagen nolecule. The opposite was true for fibronectin, for which the binding affinities were greater when fibronectin was immobilized. This was probably due to the conformational changes that occurred when fibronectin was bound to the chip. The recombinant proteins may need fibronectin chains to be in a specific fold, conformation, and/or proximity to bind with the highest affinity.

Preliminary data determined that DufA and SwpA bind a C-terminal fragment of fibronectin, that is, the Hep2-40K region (Table 29 and Figure 32). This differs from most MSCRAMMs and fibronectin binding proteins (FnBPs), which bind the N-terminal NTD-70K region (Henderson et al., 2011), to which SwpA and DufA did not bind at all (data not shown). To our knowledge, out of the many FnBPs that have been identified, only a few bind the Hep2-40K region. It would be interesting to characterize the binding specificity/affinities of SwpA and DufA to regions of fibronectin, as the Hep2-40K and NTD-70K regions of fibronectin were the only two tested in addition to full-length fibronectin. It is reasonable to think that another region of fibronectin is involved in these interactions as the affinity to immobilized full length fibronectin was greater than the binding affinity to immobilized Hep2-40K alone. Of note, bacterial FnBPs often have binding sites for multiple host ligands, and it is not uncommon for

FnBPs to preferentially bind insoluble fibronectin over soluble fibronectin (Henderson et al., 2011); we see these characteristics in SwpA and DufA.

Mutants were made by deleting the whole WxL loci individually in *E. faecium* TX82. The involvement of these loci in binding to type I collagen (seen *in vitro* with recombinant proteins) was confirmed in whole-cell binding assays where Locus A and Locus C deletion mutants were attenuated compared to the parental strain in binding to type I collagen. This was specific to type I collagen, as significant attenuation was not seen in either mutant in binding to type V collagen (Figure 34). The WxL Locus A and C deletion mutants were also tested for biofilm formation, but they were not attenuated (Figure 35). However, we cannot rule out that the WxL loci are involved in biofilm formation since there is much redundancy in contributions to biofilm and many different proteins are known to be involved (Heikens et al., 2007; Heikens et al., 2011; Hendrickx et al., 2009a; Sillanpaa et al., 2010).

Similar to lactobacilli, it seems that the *E. faecium* WxL genes are involved in response to bile salt stress and affected by salt concentrations. Transcriptomic analysis of lactobacilli found that WxL genes are regulated in response to bile, salt, and lactate stress (Siezen et al., 2006). We found that, not only do recombinant proteins undergo conformational changes under different salt concentrations (Figure 24), but salt concentrations also affect their ability to bind to each other (Figure 33). In addition, we found that the WxL Locus B mutant actually survives better than the parental strain TX82 after bile salt exposure (Figure 36).

Since the WxL proteins are found on the bacterial cell surface and have some functions potentially related to virulence, this suggested the probability that these proteins would be antigenic during infection. Indeed, we found that serum from all five of the patients with *E. faecium* endocarditis tested had antibodies to SwpA, LwpA, SwpC, and LwpC (Figure 37). On

the other hand, seven healthy volunteers were also tested. None of them contained antibodies to all four WxL proteins tested, as the endocarditis patients did, however, some contained antibodies to one of the four. Three healthy volunteers cariied antibodies to LwpC and one to LwpA while three of the healthy volunteers had no antibodies to the WxL proteins. Since *E. faecium* are commensals of the GI tract and these proteins seem to be expressed on the surface of the bacterial cell regardless of their clinical association, it appears that healthy individuals are exposured and make antibodies to these proteins as well. This finding supports the fact that WxL proteins are antigenic.

Due to the fact that antibodies to LwpA and LwpC were found in both healthy and infected patients and are the two largest proteins, we hypothesized that these two were potentially the most antigenic, and therefore might be the most effective in prophylactic immunization. We did two different experiments in which rats were immunized with a mix of LwpA and LwpC over a six week period, then challenged with TX82 in a rat endocarditis infection model via tail vein injection (Figure 38). Although in the first experiment all 8 immunized rats were protected versus the non-immunized, 2 out of 3 sham-immunized (just FCA and FICA) rats were as well. The second experiment confirmed that sham-immunized rats were protected as well as immunized rats, and therefore we could not confirm or deny the effectiveness of immunization with LwpA and LwpC. To circumvent the situation of the adjuvant resulting in a protective effect, passive immunization for protection could be used or perhaps an adjuvant that stimulates the immune system less.

[5] CONCLUDING REMARKS

The purpose of this work was to further the understanding of how an organism that was practically non–existent in the clinical setting 40 years ago could become one of the most common gram positive organisms in nosocomial infections, as well as one of the most difficult to treat. Before becoming a "no ESCKAPE" pathogen (as deemed by the IDSA), *E. faecium* was used as a probiotic, a cheese starter, and as a food additive. The first hint of a change in *E. faecium* strains was reported in a retrospective study showing an increase in beta-lactam and gentamicin resistance from 1968 to 1990 within one hospital (Grayson et al., 1991). The work presented in this dissertation has demonstrated that HA isolates have been circulating in the United States as early as 1982, where HA isolates mostly belonged to distinct MLST types different from CA isolates and were characterized by increasing ampicillin MICs (Chapter 1 Section 2.3.1 and (Galloway-Pena et al., 2009)). At first, these increases seen in ampicillin MICs were thought to have evolved from point mutations occurring in the ampicillin susceptible strains over time. However, our analysis of the *pbp5* sequences of these outbreak strains, as well as susceptible strains, showed that the nucleotide differences (~5%) between CA and HA isolates were throughout the protein and not just amino acid changes resulting in increased ampicillin resistance (Chapter 1 Section 2.3.3 and (Galloway-Pena et al., 2011). Upon extending our analysis to 3 more genes besides *pbp5* (Chapter 1 Section 2.3.3 and (Galloway-Pena et al., 2011)) and then 100 core genes (Chapter 2 Section 3.3.1 and (Galloway-Pena et al., 2012)), a 3-10% nucleotide difference was found between these genes in the HA subgroup versus the CA subgroup. The molecular clock analysis using the synonymous SNPs of the 100 core genes predicted the divergence between the HA and CA clades likely occurred hundreds of millennia ago, or longer (Chapter 2 Sections 3.3.2, 3.3.3, and (Galloway-Pena et

al., 2012)). A number of studies had previously suggested that the driving force behind the recent success of *E. faecium* in hospitals was the gain of mobile genetic elements carrying antibiotic resistance determinants, virulence and/or colonization factors. However, showing that *E. faecium* strains belong to one of two subpopulations, the HA-clade or the CA-clade, that differ at the core genome level provides evidence that acquired elements are not the only factors behind the recent success of *E. faecium*.

In addition to the differences at the core genome level seen between CA and HA isolates, some acquired genes and putative virulence genes are indeed associated with HA isolates but uncommon in CA isolates. The presence of putative virulence factors such as esp_{Efm}, and the hyl_{Efm}-containing plasmid were almost exclusively associated with HA clade strains. It was also found that the *fms* genes (putative pili or adhesins) were more often associated with HA isolates than CA isolates (Chapter 1 Section 2.3.1 and (Galloway-Pena et al., 2009)). This is further supported by the epidemiology of VRE in South America, in which the population genetics mirror that early United States isolates (Chapter 1 Section 2.3.2 and (Panesso et al., 2010)).

The differences between the CA and HA clades were further supported by the completion of the TX16 genome, providing insight into the genomic features of *E. faecium*. By analyzing the TX16 genome and comparing it to 21 draft genomes, we were able to confirm the differences between HA-clade and CA-clade isolates at the core genome level as well as the differences in the presence of IS elements, transposons, antibiotic resistance determinants, and CRISPRS (Chapter 2 Sections 3.3.5, 3.3.6, 3.3.7, and (Qin et al., 2012)). All of this information together has helped to further the understanding as to what factors most likely contributed to the emergence and success of this organism in the hospital environment.

These studies also provide insight into the question, "what is the difference between the epidemiology of VREfm seen in the United States versus Europe"? As discussed in the Background Section, there is a discrepancy between the epidemiology of VREfm seen in the United States and Europe. VREfm are, and have been for the past 20 years, a common cause of healthcare associated infections in the United States, yet VREfm are not found in the community in the United States (Coque et al., 1996). On the other hand, VREfm are a rare (albeit now an increasing) cause of infection in the European Union, yet they are commonly found in fecal samples of animals and healthy individuals (Bonten et al., 2001; Willems et al., 2000; Willems et al., 2005). We can now conclude that the differences between these two geographical areas were due to the differences between CA and HA strains as well as a contribution of ampicillin resistance. It seems from our data that, in the United States, *E. faecium* strains containing *pbp5*-R allele of the HA clade acquired vancomycin resistance. Conversely, in the EU, *E. faecium* strains containing *pbp5*-S of the CA clade were the first to acquire vancomycin resistance which was then perpetuated and increased due to the use of avoparcin in animal husbandry. Therefore, it seems the EU VREfm were not as likely to survive in hospitals as the United States isolates due to their lack of multiresistance, putative virulence factors, and the advantage that comes with being a HA clade strain.

After sequencing three genomes (TX16, TX1330, and TX82) within our own laboratory, we furthered our analysis by looking for novel pili, adhesins, and surface proteins. The discovery of an atypical LPxTG-like protein (*fms6*) led to the finding of three apparent operons in *E. faecium* encoding proteins with the recently identified WxL motif. Further *in vitro* and *in vivo* characterization showed complex proteins that localize to the surface of the bacterial cell and bind each other, possibly making a novel cell surface complex. The protein

conformation, folding, and binding affinities seem to be affected by salt concentration. Analysis also suggests that these operons bind numerous host extracellular matrix tissues and could be involved in bile stress, making these operons multifunctional. In addition, the fact that these proteins are highly antigenic may give hope for novel candidates for immunization, which will help us to bypass the lack of effective antibiotics available for enterococcal infections.

The study of the WxL proteins opens many doors to future experiments and analyses to further understand these complex proteins. One area would be to understand if the function of the WxL proteins actually influences virulence. A second line of investigation would be to compare the expression of these proteins between community and clinical strains at the transcript level. Furthermore, one could determine the *in vitro* versus *in vivo* expression, i.e., are these proteins expressed more during infection. A third line of investigation would be to determine which residues of the WxL domain are important for cell surface localization and which regions are responsible for interactions with host tissues. In addition, one could further investigate if the proteins within the same operon actually interact and/or colocalize and create novel cell surface complexes.

In conclusion, although *E. faecium* is now an important hospital-associated pathogen, research on how this historically commensal organism initiates disease and causes infection within the host is deficient when compared to other Gram positive pathogens. Determining the specific genomic features that contribute to pathogenicity and persistence in the hospital environment is key to understanding *E. faecium*'s transition between the community and hospital niches. This includes both genes that are unique or enriched in the HA clade, as well as core genomic differences. Furthermore, assessment of the importance of putative surface proteins (i.e those with the "WxL" motif) in virulence, protein-protein interactions, and the

expression of the genes within the host may enhance the very limited knowledge of the mechanisms for pathogenicity of this organism, and possibly provide alternatives for the prevention and/or treatment of enterococcal infections.

BIBLIOGRAPHY

Achtman, M., G. Morelli, P. Zhu, T. Wirth, I. Diehl, B. Kusecek, A.J. Vogler, D.M. Wagner, C.J. Allender, W.R. Easterday, V. Chenal-Francisque, P. Worsham, N.R. Thomson, J. Parkhill, L.E. Lindler, E. Carniel, and P. Keim. 2004. Microevolution and history of the plague bacillus, *Yersinia pestis*. *Proc Natl Acad Sci U S A* 101(51):17837-17842.

Arias, C.A., and B.E. Murray. 2008. Emergence and management of drug-resistant enterococcal infections. *Expert Rev Anti Infect Ther* 6(5):637-655.

Arias, C.A., and B.E. Murray. 2012. The rise of the Enterococcus: beyond vancomycin resistance. *Nat Rev Microbiol* 10(4):266-278.

Arias, C.A., D. Panesso, K.V. Singh, L.B. Rice, and B.E. Murray. 2009. Cotransfer of antibiotic resistance genes and a hylEfm-containing virulence plasmid in Enterococcus faecium. *Antimicrob Agents Chemother* 53(10):4240-4246.

Arias, C.A., J. Reyes, M. Zuniga, L. Cortes, C. Cruz, C.L. Rico, and D. Panesso. 2003. Multicentre surveillance of antimicrobial resistance in enterococci and staphylococci from Colombian hospitals, 2001-2002. *J Antimicrob Chemother* 51(1):59-68.

Arthur, M., C. Molinas, F. Depardieu, and P. Courvalin. 1993. Characterization of *Tn1546*, a *Tn3*-related transposon conferring glycopeptide resistance by synthesis of depsipeptide peptidoglycan precursors in *Enterococcus faecium* BM4147. *J Bacteriol* 175(1):117-127.

Bland, C., T.L. Ramsey, F. Sabree, M. Lowe, K. Brown, N.C. Kyrpides, and P. Hugenholtz. 2007. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* 8:209.

Bonten, M.J., R. Willems, and R.A. Weinstein. 2001. Vancomycin-resistant enterococci: why are they here, and where do they come from? *Lancet Infect Dis* 1(5):314-325.

Bourgogne, A., D.A. Garsin, X. Qin, K.V. Singh, J. Sillanpaa, S. Yerrapragada, Y. Ding, S. Dugan-Rocha, C. Buhay, H. Shen, G. Chen, G. Williams, D. Muzny, A. Maadani, K.A. Fox, J. Gioia, L. Chen, Y. Shang, C.A. Arias, S.R. Nallapareddy, M. Zhao, V.P. Prakash, S. Chowdhury, H. Jiang, R.A. Gibbs, B.E. Murray, S.K. Highlander, and G.M. Weinstock. 2008. Large scale variation in *Enterococcus faecalis* illustrated by the genome analysis of strain OG1RF. *Genome Biol* 9(7):R110.

Boyce, J.M., S.M. Opal, G. Potter-Bynoe, R.G. LaForge, M.J. Zervos, G. Furtado, G. Victor, and A.A. Medeiros. 1992. Emergence and nosocomial transmission of ampicillin-resistant enterococci. *Antimicrob Agents Chemother* 36(5):1032-1039.

Brinster, S., S. Furlan, and P. Serror. 2007a. C-terminal WxL domain mediates cell wall binding in *Enterococcus faecalis* and other gram-positive bacteria. *J Bacteriol* 189(4):1244-1253.

Brinster, S., B. Posteraro, H. Bierne, A. Alberti, S. Makhzami, M. Sanguinetti, and P. Serror. 2007b. Enterococcal leucine-rich repeat-containing protein involved in virulence and host inflammatory response. *Infect Immun* 75(9):4463-4471.

Bruen, T.C., H. Philippe, and D. Bryant. 2006. A simple and robust statistical test for detecting the presence of recombination. *Genetics* 172(4):2665-2681.

Burdett, V., J. Inamine, and S. Rajagopalan. 1982. Heterogeneity of tetracycline resistance determinants in Streptococcus. *Journal of bacteriology* 149(3):995-1004.

Bush, L.M., J. Calmon, C.L. Cherney, M. Wendeler, P. Pitsakis, J. Poupard, M.E. Levison, and C.C. Johnson. 1989. High-level penicillin resistance among isolates of enterococci. Implications for treatment of enterococcal infections. *Ann Intern Med* 110(7):515-520.

Carias, L.L., S.D. Rudin, C.J. Donskey, and L.B. Rice. 1998. Genetic linkage and cotransfer of a novel, *vanB*-containing transposon (Tn*5382*) and a low-affinity penicillin-binding protein 5 gene in a clinical vancomycin-resistant *Enterococcus faecium* isolate. *J Bacteriol* 180(17):4426-4434.

Chen, L., J. Yang, J. Yu, Z. Yao, L. Sun, Y. Shen, and Q. Jin. 2005. VFDB: a reference database for bacterial virulence factors. *Nucleic acids research* 33 (Database issue):D325-328.

Chenna, R., H. Sugawara, T. Koike, R. Lopez, T.J. Gibson, D.G. Higgins, and J.D. Thompson. 2003. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* 31(13):3497-3500.

Choudhury, T., K.V. Singh, J. Sillanpaa, S.R. Nallapareddy, and B.E. Murray. 2011. Importance of two *Enterococcus faecium* loci encoding Gls-like proteins for in vitro bile salts stress response and virulence. *J Infect Dis* 203(8):1147-1154.

Coque, T.M., J.E. Patterson, J.M. Steckelberg, and B.E. Murray. 1995. Incidence of hemolysin, gelatinase, and aggregation substance among enterococci isolated from patients with endocarditis and other infections and from feces of hospitalized and community-based persons. *J Infect Dis* 171(5):1223-1229.

Coque, T.M., J.F. Tomayko, S.C. Ricke, P.C. Okhyusen, and B.E. Murray. 1996. Vancomycin-resistant enterococci from nosocomial, community, and animal sources in the United States. *Antimicrob Agents Chemother* 40(11):2605-2609.

Coudron, P.E., C.G. Mayhall, R.R. Facklam, A.C. Spadora, V.A. Lamb, M.R. Lybrand, and H.P. Dalton. 1984. *Streptococcus faecium* outbreak in a neonatal intensive care unit. *J Clin Microbiol* 20(6):1044-1048.

Dalla Costa, L.M., D.C. Souza, L.T. Martins, R.C. Zanella, M.C. Brandilone, S. Bokermann, H.S. Sader, and H.A. Souza. 1998. Vancomycin-Resistant *Enterococcus faecium*: First Case in Brazil. *Braz J Infect Dis* 2(3):160-163.

Didelot, X., and D. Falush. 2007. Inference of bacterial microevolution using multilocus sequence data. *Genetics* 175(3):1251-1266.

Dutka-Malen, S., S. Evers, and P. Courvalin. 1995. Detection of glycopeptide resistance genotypes and identification to the species level of clinically relevant enterococci by PCR. *J Clin Microbiol* 33(1):24-27.

Eliopoulos, G.M., C. Wennersten, and R.C. Moellering, Jr. 1982. Resistance to beta-lactam antibiotics in *Streptococcus faecium*. *Antimicrob Agents Chemother* 22(2):295-301.

Eliopoulos, G.M., C. Wennersten, S. Zighelboim-Daum, E. Reiszner, D. Goldmann, and R.C. Moellering, Jr. 1988. High-level resistance to gentamicin in clinical isolates of *Streptococcus (Enterococcus) faecium*. *Antimicrob Agents Chemother* 32(10):1528-1532.

Fisher, K., and C. Phillips. 2009. The ecology, epidemiology and virulence of *Enterococcus*. *Microbiology* 155(Pt 6):1749-1757.

Fontana, R., M. Aldegheri, M. Ligozzi, H. Lopez, A. Sucari, and G. Satta. 1994. Overproduction of a low-affinity penicillin-binding protein and high-level ampicillin resistance in *Enterococcus faecium*. *Antimicrob Agents Chemother* 38(9):1980-1983.

Fontana, R., A. Grossato, L. Rossi, Y.R. Cheng, and G. Satta. 1985. Transition from resistance to hypersusceptibility to beta-lactam antibiotics associated with loss of a low-affinity penicillin-binding protein in a *Streptococcus faecium* mutant highly resistant to penicillin. *Antimicrob Agents Chemother* 28(5):678-683.

Foster, J.T., S.M. Beckstrom-Sternberg, T. Pearson, J.S. Beckstrom-Sternberg, P.S. Chain, F.F. Roberto, J. Hnath, T. Brettin, and P. Keim. 2009. Whole-genome-based phylogeny and divergence of the genus Brucella. *J Bacteriol* 191(8):2864-2870.

Foulquie Moreno, M.R., P. Sarantinopoulos, E. Tsakalidou, and L. De Vuyst. 2006. The role and application of enterococci in food and health. *Int J Food Microbiol* 106(1):1-24.

Galloway-Pena, J., J.H. Roh, M. Latorre, X. Qin, and B.E. Murray. 2012. Genomic and SNP Analyses Demonstrate a Distant Separation of the Hospital and Community-Associated Clades of *Enterococcus faecium*. *PLoS One* 7(1):e30187.

Galloway-Pena, J.R., S.R. Nallapareddy, C.A. Arias, G.M. Eliopoulos, and B.E. Murray. 2009. Analysis of clonality and antibiotic resistance among early clinical isolates of *Enterococcus faecium* in the United States. *J Infect Dis* 200(10):1566-1573.

Galloway-Pena, J.R., L.B. Rice, andB.E. Murray. 2011. Analysis of PBP5 of early U.S. isolates of *Enterococcus faecium*: sequence variation alone does not explain increasing ampicillin resistance over time. *Antimicrob Agents Chemother* 55(7):3272-3277.

Gao, P., K.L. Pinkston, S.R. Nallapareddy, A. van Hoof, B.E. Murray, and B.R. Harvey. 2010. *Enterococcus faecalis rnjB* is required for pilin gene expression and biofilm formation. *J Bacteriol* 192(20):5489-5498.

Grady, R., and F. Hayes. 2003. Axe-Txe, a broad-spectrum proteic toxin-antitoxin system specified by a multidrug-resistant, clinical isolate of *Enterococcus faecium*. *Mol Microbiol* 47(5):1419-1432.

Grayson, M.L., G.M. Eliopoulos, C.B. Wennersten, K.L. Ruoff, P.C. De Girolami, M.J. Ferraro, and R.C. Moellering, Jr. 1991. Increasing resistance to beta-lactam antibiotics among clinical isolates of *Enterococcus faecium*: a 22-year review at one institution. *Antimicrob Agents Chemother* 35(11):2180-2184.

Guindon, S., J.F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology* 59(3):307-321.

Handwerger, S., B. Raucher, D. Altarac, J. Monka, S. Marchione, K.V. Singh, B.E. Murray, J. Wolff, and B. Walters. 1993. Nosocomial outbreak due to *Enterococcus faecium* highly resistant to vancomycin, penicillin, and gentamicin. *Clin Infect Dis* 16(6):750-755.

Heikens, E., M.J. Bonten, and R.J. Willems. 2007. Enterococcal surface protein Esp is important for biofilm formation of *Enterococcus faecium* E1162. *J Bacteriol* 189(22):8233-8240.

Heikens, E., M. Leendertse, L.M. Wijnands, M. van Luit-Asbroek, M.J. Bonten, T. van der Poll, and R.J. Willems. 2009. Enterococcal surface protein Esp is not essential for cell adhesion and intestinal colonization of *Enterococcus faecium* in mice. *BMC Microbiol* 9:19.

Heikens, E., K.V. Singh, K.D. Jacques-Palaz, M. van Luit-Asbroek, E.A. Oostdijk, M.J. Bonten, B.E. Murray, and R.J. Willems. 2011. Contribution of the enterococcal surface

protein Esp to pathogenesis of *Enterococcus faecium* endocarditis. *Microbes Infect* 13(14-15):1185-1190.

Henderson, B., S. Nair, J. Pallas, and M.A. Williams. 2011. Fibronectin: a multidomain host adhesin targeted by bacterial fibronectin-binding proteins. *FEMS Microbiol Rev* 35(1):147-200.

Hendrickx, A.P., M.J. Bonten, M. van Luit-Asbroek, C.M. Schapendonk, A.H. Kragten, and R.J. Willems. 2008. Expression of two distinct types of pili by a hospital-acquired *Enterococcus faecium* isolate. *Microbiology* 154(Pt 10):3212-3223.

Hendrickx, A.P., M. van Luit-Asbroek, C.M. Schapendonk, W.J. van Wamel, J.C. Braat, L.M. Wijnands, M.J. Bonten, and R.J. Willems. 2009a. SgrA, a nidogen-binding LPXTG surface adhesin implicated in biofilm formation, and EcbA, a collagen binding MSCRAMM, are two novel adhesins of hospital-acquired *Enterococcus faecium*. *Infect Immun* 77(11):5097-5106.

Hendrickx, A.P., W.J. van Wamel, G. Posthuma, M.J. Bonten, and R.J. Willems. 2007. Five genes encoding surface-exposed LPXTG proteins are enriched in hospital-adapted *Enterococcus faecium* clonal complex 17 isolates. *J Bacteriol* 189(22):8321-8332.

Hendrickx, A.P., R.J. Willems, M.J. Bonten, and W. van Schaik. 2009b. LPxTG surface proteins of enterococci. *Trends Microbiol* 17(9):423-430.

Hidron, A.I., J.R. Edwards, J. Patel, T.C. Horan, D.M. Sievert, D.A. Pollock, and S.K. Fridkin. 2008. NHSN annual update: antimicrobial-resistant pathogens associated with healthcare-associated infections: annual summary of data reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2006-2007. *Infect Control Hosp Epidemiol* 29(11):996-1011.

Homan, W.L., D. Tribe, S. Poznanski, M. Li, G. Hogg, E. Spalburg, J.D. Van Embden, and R.J. Willems. 2002. Multilocus sequence typing scheme for *Enterococcus faecium*. *J Clin Microbiol* 40(6):1963-1971.

Jureen, R., S. Harthug, S. Sornes, A. Digranes, R.J. Willems, and N. Langeland. 2004a. Comparative analysis of amplified fragment length polymorphism and pulsed field gel electrophoresis in a hospital outbreak and subsequent endemicity of ampicillin-resistant *Enterococcus faecium*. *FEMS Immunol Med Microbiol* 40(1):33-39.

Jureen, R., S.C. Mohn, S. Harthug, L. Haarr, and N. Langeland. 2004b. Role of penicillin-binding protein 5 C-terminal amino acid substitutions in conferring ampicillin resistance in Norwegian clinical strains of *Enterococcus faecium*. *Apmis* 112(4-5):291-298.

Katoh, K., K. Misawa, K. Kuma, and T. Miyata. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research* 30(14):3059-3066.

Khan, M.A., J.B. Northwood, R.G. Loor, A.T. Tholen, E. Riera, M. Falcon, A. van Belkum, M. van Westreenen, and J.P. Hays. 2009. High prevalence of ST-78 infection-associated vancomycin-resistant *Enterococcus faecium* from hospitals in Asuncion, Paraguay. *Clin Microbiol Infect*.

Kim, D.S., K.V. Singh, S.R. Nallapareddy, X. Qin, D. Panesso, C.A. Arias, and B.E. Murray. 2010. The *fms21* (*pilA*)-*fms20* locus encoding one of four distinct pili of *Enterococcus faecium* is harboured on a large transferable plasmid associated with gut colonization and virulence. *J Med Microbiol* 59(Pt 4):505-507.

Klare, I., A.C. Rodloff, J. Wagner, W. Witte, and R. Hakenbeck. 1992. Overproduction of a penicillin-binding protein is not the only mechanism of penicillin resistance in *Enterococcus faecium*. *Antimicrob Agents Chemother* 36(4):783-787.

Klein, G. 2003. Taxonomy, ecology and antibiotic resistance of enterococci from food and the gastro-intestinal tract. *Int J Food Microbiol* 88(2-3):123-131.

Klibi, N., Y. Saenz, M. Zarazaga, K. Ben Slama, A. Masmoudi, F. Ruiz-Larrea, A. Boudabous, and C. Torres. 2008. Polymorphism in *pbp5* gene detected in clinical *Enterococcus faecium* strains with different ampicillin MICs from a Tunisian hospital. *J Chemother* 20(4):436-440.

Korten, V., W.M. Huang, and B.E. Murray. 1994. Analysis by PCR and direct DNA sequencing of *gyrA* mutations associated with fluoroquinolone resistance in *Enterococcus faecalis*. *Antimicrob Agents Chemother* 38(9):2091-2094.

Kristich, C.J., J.R. Chandler, and G.M. Dunny. 2007. Development of a host-genotype-independent counterselectable marker and a high-frequency conjugative delivery system and their use in genetic analysis of Enterococcus faecalis. *Plasmid* 57(2):131-144.

Kumar, S., M. Nei, J. Dudley, and K. Tamura. 2008. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform* 9(4):299-306.

Kuo, C.H., and H. Ochman. 2009. Inferring clocks when lacking rocks: the variable rates of molecular evolution in bacteria. *Biol Direct* 4:35.

Lam, M.M., T. Seemann, D.M. Bulach, S.L. Gladman, H. Chen, V. Haring, R.J. Moore, S. Ballard, M.L. Grayson, P.D. Johnson, B.P. Howden, and T.P. Stinear. 2012. Comparative Analysis of the First Complete *Enterococcus faecium* Genome. *J Bacteriol* 194(9):2334-2341.

Larkin, M.A., G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, and D.G. Higgins. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23(21):2947-2948.

Leavis, H., J. Top, N. Shankar, K. Borgen, M. Bonten, J. van Embden, and R.J. Willems. 2004. A novel putative enterococcal pathogenicity island linked to the *esp* virulence gene of *Enterococcus faecium* and associated with epidemicity. *J Bacteriol* 186(3):672-682.

Leavis, H.L., M.J. Bonten, and R.J. Willems. 2006a. Identification of high-risk enterococcal clonal complexes: global dispersion and antibiotic resistance. *Curr Opin Microbiol* 9(5):454-460.

Leavis, H.L., R.J. Willems, J. Top, and M.J. Bonten. 2006b. High-level ciprofloxacin resistance from point mutations in *gyrA* and *parC* confined to global hospital-adapted clonal lineage CC17 of *Enterococcus faecium*. *J Clin Microbiol* 44(3):1059-1064.

Leavis, H.L., R.J. Willems, W.J. van Wamel, F.H. Schuren, M.P. Caspers, and M.J. Bonten. 2007. Insertion sequence-driven diversification creates a globally dispersed emerging multiresistant subspecies of *E. faecium*. *PLoS Pathog* 3(1):e7.

Leendertse, M., E. Heikens, L.M. Wijnands, M. van Luit-Asbroek, G.J. Teske, J.J. Roelofs, M.J. Bonten, T. van der Poll, and R.J. Willems. 2009. Enterococcal surface protein transiently aggravates *Enterococcus faecium*-induced urinary tract infection in mice. *J Infect Dis* 200(7):1162-1165.

Lenski, R.E., C.L. Winkworth, and M.A. Riley. 2003. Rates of DNA sequence evolution in experimental populations of *Escherichia coli* during 20,000 generations. *J Mol Evol* 56(4):498-508.

Lester, C.H., D. Sandvang, S.S. Olsen, H.C. Schonheyder, J.O. Jarlov, J. Bangsborg, D.S. Hansen, T.G. Jensen, N. Frimodt-Moller, and A.M. Hammerum. 2008. Emergence of ampicillin-resistant *Enterococcus faecium* in Danish hospitals. *J Antimicrob Chemother* 62(6):1203-1206.

Letunic, I., and P. Bork. 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23(1):127-128.

Li, L., C.J. Stoeckert, Jr., and D.S. Roos. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 13(9):2178-2189.

Ligozzi, M., F. Pittaluga, and R. Fontana. 1996. Modification of penicillin-binding protein 5 associated with high-level ampicillin resistance in *Enterococcus faecium*. *Antimicrob Agents Chemother* 40(2):354-357.

Lopez, M., J.C. Hormazabal, A. Maldonado, G. Saavedra, F. Baquero, J. Silva, C. Torres, and R. del Campo. 2009. Clonal dissemination of *Enterococcus faecalis* ST201 and *Enterococcus faecium* CC17-ST64 containing Tn*5382-vanB2* among 16 hospitals in Chile. *Clin Microbiol Infect* 15(6):586-588.

Mandlik, A., A. Das, and H. Ton-That. 2008. The molecular switch that activates the cell wall anchoring step of pilus assembly in gram-positive bacteria. *Proc Natl Acad Sci U S A* 105(37):14147-14152.

Marin, M.E., J.R. Mera, R.C. Arduino, A.P. Correa, T.M. Coque, D. Stamboulian, and B.E. Murray. 1998. First report of vancomycin-resistant *Enterococcus faecium* isolated in Argentina. *Clin Infect Dis* 26(1):235-236.

Miranda, A.G., K.V. Singh, and B.E. Murray. 1991. DNA fingerprinting of *Enterococcus faecium* by pulsed-field gel electrophoresis may be a useful epidemiologic tool. *J Clin Microbiol* 29(12):2752-2757.

Moet, G.J., R.N. Jones, D.J. Biedenbach, M.G. Stilwell, and T.R. Fritsche. 2007. Contemporary causes of skin and soft tissue infections in North America, Latin America, and Europe: report from the SENTRY Antimicrobial Surveillance Program (1998-2004). *Diagn Microbiol Infect Dis* 57(1):7-13.

Mohamed, J.A., W. Huang, S.R. Nallapareddy, F. Teng, and B.E. Murray. 2004. Influence of origin of isolates, especially endocarditis isolates, and various genes on biofilm formation by *Enterococcus faecalis*. *Infect Immun* 72(6):3658-3663.

Murphy, E., L. Huwyler, and C. de Freire Bastos Mdo. 1985. Transposon *Tn554*: complete nucleotide sequence and isolation of transposition-defective and antibiotic-sensitive mutants. *Embo J* 4(12):3357-3365.

Murray, B.E. 1990. The life and times of the Enterococcus. *Clin Microbiol Rev* 3(1):46-65.

Murray, B.E. 1997. Vancomycin-resistant enterococci. *Am J Med* 102(3):284-293.

Nallapareddy, S.R., X. Qin, G.M. Weinstock, M. Hook, and B.E. Murray. 2000. *Enterococcus faecalis* adhesin, *ace*, mediates attachment to extracellular matrix proteins collagen type IV and laminin as well as collagen type I. *Infect Immun* 68(9):5218-5224.

Nallapareddy, S.R., K.V. Singh, and B.E. Murray. 2006. Construction of improved temperature-sensitive and mobilizable vectors and their use for constructing mutations in the adhesin-encoding acm gene of poorly transformable clinical *Enterococcus faecium* strains. *Appl Environ Microbiol* 72(1):334-345.

Nallapareddy, S.R., K.V. Singh, and B.E. Murray. 2008a. Contribution of the collagen adhesin Acm to pathogenesis of *Enterococcus faecium* in experimental endocarditis. *Infect Immun* 76(9):4120-4128.

Nallapareddy, S.R., K.V. Singh, P.C. Okhuysen, and B.E. Murray. 2008b. A functional collagen adhesin gene, *acm*, in clinical isolates of *Enterococcus faecium* correlates with the recent success of this emerging nosocomial pathogen. *Infect Immun* 76(9):4110-4119.

Nallapareddy, S.R., G.M. Weinstock, and B.E. Murray. 2003. Clinical isolates of *Enterococcus faecium* exhibit strain-specific collagen binding mediated by Acm, a new member of the MSCRAMM family. *Mol Microbiol* 47(6):1733-1747.

Ochman, H., S. Elwyn, and N.A. Moran. 1999. Calibrating bacterial evolution. *Proc Natl Acad Sci U S A* 96(22):12638-12643.

Ochman, H., and A.C. Wilson. 1987. Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *J Mol Evol* 26(1-2):74-86.

Palmer, K.L., K. Carniol, J.M. Manson, D. Heiman, T. Shea, S. Young, Q. Zeng, D. Gevers, M. Feldgarden, B. Birren, and M.S. Gilmore. 2010. High-quality draft genome sequences of 28 Enterococcus sp. isolates. *J Bacteriol* 192(9):2469-2470.

Palmer, K.L., andM.S. Gilmore. 2010. Multidrug-resistant enterococci lack CRISPR-cas. *MBio* 1(4).

Palmer, K.L., P. Godfrey, A. Griggs, V.N. Kos, J. Zucker, C. Desjardins, G. Cerqueira, D. Gevers, S. Walker, J. Wortman, M. Feldgarden, B. Haas, B. Birren, and M.S. Gilmore. 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):e00318-00311.

Panesso, D., M.C. Montealegre, S. Rincon, M.F. Mojica, L.B. Rice, K.V. Singh, B.E. Murray, and C.A. Arias. 2011. The *hyl*Efm gene in pHylEfm of *Enterococcus faecium* is not required in pathogenesis of murine peritonitis. *BMC Microbiol* 11(1):20.

Panesso, D., J. Reyes, S. Rincon, L. Diaz, J. Galloway-Pena, J. Zurita, C. Carrillo, A. Merentes, M. Guzman, J.A. Adachi, B.E. Murray, and C.A. Arias. 2010. Molecular epidemiology of vancomycin-resistant *Enterococcus faecium*: a prospective, multicenter study in South American hospitals. *J Clin Microbiol* 48(5):1562-1569.

Pearson, T., P. Giffard, S. Beckstrom-Sternberg, R. Auerbach, H. Hornstra, A. Tuanyok, E.P. Price, M.B. Glass, B. Leadem, J.S. Beckstrom-Sternberg, G.J. Allan, J.T. Foster, D.M. Wagner, R.T. Okinaka, S.H. Sim, O. Pearson, Z. Wu, J. Chang, R. Kaul, A.R. Hoffmaster, T.S. Brettin, R.A. Robison, M. Mayo, J.E. Gee, P. Tan, B.J. Currie, and P. Keim. 2009. Phylogeographic reconstruction of a bacterial species with high levels of lateral gene transfer. *BMC Biol* 7:78.

Pinkston, K.L., P. Gao, D. Diaz-Garcia, J. Sillanpaa, S.R. Nallapareddy, B.E. Murray, and B.R. Harvey. 2011. The Fsr quorum-sensing system of *Enterococcus faecalis* modulates surface display of the collagen-binding MSCRAMM Ace through regulation of *gelE*. *J Bacteriol* 193(17):4317-4325.

Poeta, P., D. Costa, G. Igrejas, Y. Saenz, M. Zarazaga, J. Rodrigues, and C. Torres. 2007. Polymorphisms of the *pbp5* gene and correlation with ampicillin resistance in *Enterococcus faecium* isolates of animal origin. *J Med Microbiol* 56(Pt 2):236-240.

Qin, X., J.R. Galloway-Pena, J. Sillanpaa, J.H. Roh, S.R. Nallapareddy, S. Chowdhury, A. Bourgogne, T. Choudhury, D.M. Muzny, C.J. Buhay, Y. Ding, S. Dugan-Rocha, W. Liu, C. Kovar, E. Sodergren, S. Highlander, J.F. Petrosino, K.C. Worley, R.A. Gibbs, G.M. Weinstock, and B.E. Murray. 2012. Complete genome sequence of *Enterococcus faecium* strain TX16 and comparative genomic analysis of *Enterococcus faecium* genomes. *BMC Microbiol* 12:135.

Rice, L.B., S. Bellais, L.L. Carias, R. Hutton-Thomas, R.A. Bonomo, P. Caspers, M.G. Page, and L. Gutmann. 2004. Impact of specific *pbp5* mutations on expression of beta-lactam resistance in *Enterococcus faecium*. *Antimicrob Agents Chemother* 48(8):3028-3032.

Rice, L.B., L. Carias, S. Rudin, C. Vael, H. Goossens, C. Konstabel, I. Klare, S.R. Nallapareddy, W. Huang, and B.E. Murray. 2003. A potential virulence gene, hyl_{Efm}, predominates in *Enterococcus faecium* of clinical origin. *J Infect Dis* 187(3):508-512.

Rice, L.B., L.L. Carias, R. Hutton-Thomas, F. Sifaoui, L. Gutmann, and S.D. Rudin. 2001. Penicillin-binding protein 5 and expression of ampicillin resistance in *Enterococcus faecium*. *Antimicrob Agents Chemother* 45(5):1480-1486.

Rice, L.B., L.L. Carias, S. Rudin, V. Lakticova, A. Wood, and R. Hutton-Thomas. 2005. *Enterococcus faecium* low-affinity *pbp5* is a transferable determinant. *Antimicrob Agents Chemother* 49(12):5007-5012.

Rice, L.B., V. Lakticova, L.L. Carias, S. Rudin, R. Hutton, and S.H. Marshall. 2009. Transferable capacity for gastrointestinal colonization in *Enterococcus faecium* in a mouse model. *J Infect Dis* 199(3):342-349.

Rybkine, T., J.L. Mainardi, W. Sougakoff, E. Collatz, and L. Gutmann. 1998. Penicillin-binding protein 5 sequence alterations in clinical isolates of *Enterococcus faecium* with different levels of beta-lactam resistance. *J Infect Dis* 178(1):159-163.

Sapico, F.L., H.N. Canawati, V.J. Ginunas, D.S. Gilmore, J.Z. Montgomerie, W.J. Tuddenham, and R.R. Facklam. 1989. Enterococci highly resistant to penicillin and ampicillin: an emerging clinical problem? *J Clin Microbiol* 27(9):2091-2095.

Sauvage, E., F. Kerff, E. Fonze, R. Herman, B. Schoot, J.P. Marquette, Y. Taburet, D. Prevost, J. Dumas, G. Leonard, P. Stefanic, J. Coyette, and P. Charlier. 2002. The 2.4-A crystal structure of the penicillin-resistant penicillin-binding protein PBP5fm from *Enterococcus faecium* in complex with benzylpenicillin. *Cell Mol Life Sci* 59(7):1223-1232.

Schachtsiek, M., W.P. Hammes, and C. Hertel. 2004. Characterization of *Lactobacillus coryniformis* DSM 20001T surface protein Cpf mediating coaggregation with and aggregation among pathogens. *Appl Environ Microbiol* 70(12):7078-7085.

Schwarz, F.V., V. Perreten, and M. Teuber. 2001. Sequence of the 50-kb conjugative multiresistance plasmid pRE25 from *Enterococcus faecalis* RE25. *Plasmid* 46(3):170-187.

Siezen, R., J. Boekhorst, L. Muscariello, D. Molenaar, B. Renckens, and M. Kleerebezem. 2006. *Lactobacillus plantarum* gene clusters encoding putative cell-surface protein complexes for carbohydrate utilization are conserved in specific gram-positive bacteria. *BMC Genomics* 7:126.

Sifaoui, F., M. Arthur, L. Rice, and L. Gutmann. 2001. Role of penicillin-binding protein 5 in expression of ampicillin resistance and peptidoglycan structure in *Enterococcus faecium*. *Antimicrob Agents Chemother* 45(9):2594-2597.

Sillanpaa, J., S.R. Nallapareddy, V.P. Prakash, X. Qin, M. Hook, G.M. Weinstock, and B.E. Murray. 2008. Identification and phenotypic characterization of a second collagen adhesin, Scm, and genome-based identification and analysis of 13 other predicted MSCRAMMs, including four distinct pilus loci, in *Enterococcus faecium*. *Microbiology* 154(Pt 10):3199-3211.

Sillanpaa, J., S.R. Nallapareddy, K.V. Singh, V.P. Prakash, T. Fothergill, H. Ton-That, and B.E. Murray. 2010. Characterization of the *ebp(fm)* pilus-encoding operon of *Enterococcus faecium* and its role in biofilm formation and virulence in a murine model of urinary tract infection. *Virulence* 1(4):236-246.

Sillanpaa, J., V.P. Prakash, S.R. Nallapareddy, and B.E. Murray. 2009a. Distribution of genes encoding MSCRAMMs and Pili in clinical and natural populations of *Enterococcus faecium*. *J Clin Microbiol* 47(4):896-901.

Sillanpaa, J., V.P. Prakash, S.R. Nallapareddy, and B.E. Murray. 2009b. Distribution of genes encoding MSCRAMMs and pili in clinical and natural populations of *Enterococcus faecium*. *J Clin Microbiol*.

Singh, K.V., T.M. Coque, and B.E. Murray. 1996. In vitro activity of the trinem sanfetrinem (GV104326) against gram-positive organisms. *Antimicrob Agents Chemother* 40(9):2142-2146.

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

Singh, K.V., S.R. Nallapareddy, J. Sillanpaa, and B.E. Murray. 2012. Importance of the collagen adhesin *ace* in pathogenesis and protection against *Enterococcus faecalis* experimental endocarditis. *PLoS Pathog* 6(1):e1000716.

Sneath, P.H., and R.R. Sokal. 1962. Numerical taxonomy. *Nature* 193:855-860.

Spiegel, C.A. 1988. Laboratory detection of high-level aminoglycoside-aminocyclitol resistance in Enterococcus spp. *J Clin Microbiol* 26(11):2270-2274.

Stratton, C.W., C. Liu, H.B. Ratner, and L.S. Weeks. 1987. Bactericidal activity of deptomycin (LY146032) compared with those of ciprofloxacin, vancomycin, and ampicillin against enterococci as determined by kill-kinetic studies. *Antimicrob Agents Chemother* 31(7):1014-1016.

Suzuki, H., T. Lefebure, M.J. Hubisz, P. Pavinski Bitar, P. Lang, A. Siepel, and M.J. Stanhope. 2011. Comparative genomic analysis of the *Streptococcus dysgalactiae* species group: gene content, molecular adaptation, and promoter evolution. *Genome Biol Evol* 3:168-185.

Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24(8):1596-1599.

Tang, J., W.P. Hanage, C. Fraser, and J. Corander. 2009. Identifying currents in the gene pool for bacterial populations using an integrative approach. *PLoS Comput Biol* 5(8):e1000455.

Teng, F., E.C. Nannini, and B.E. Murray. 2005. Importance of *gls24* in virulence and stress response of *Enterococcus faecalis* and use of the Gls24 protein as a possible immunotherapy target. *J Infect Dis* 191(3):472-480.

Thompson, J.D., D.G. Higgins, and T.J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22(22):4673-4680.

Top, J., R. Willems, H. Blok, M. de Regt, K. Jalink, A. Troelstra, B. Goorhuis, and M. Bonten. 2007. Ecological replacement of *Enterococcus faecalis* by multiresistant clonal complex 17 *Enterococcus faecium*. *Clin Microbiol Infect* 13(3):316-319.

Top, J., R. Willems, and M. Bonten. 2008. Emergence of CC17 *Enterococcus faecium*: from commensal to hospital-adapted pathogen. *FEMS Immunol Med Microbiol* 52(3):297-308.

Turner, K.M., W.P. Hanage, C. Fraser, T.R. Connor, and B.G. Spratt. 2007. Assessing the reliability of eBURST using simulated populations with known ancestry. *BMC Microbiol* 7:30.

Uttley, A.H., C.H. Collins, J. Naidoo, and R.C. George. 1988. Vancomycin-resistant enterococci. *Lancet* 1(8575-6):57-58.

Van Ert, M.N., W.R. Easterday, L.Y. Huynh, R.T. Okinaka, M.E. Hugh-Jones, J. Ravel, S.R. Zanecki, T. Pearson, T.S. Simonson, J.M. U'Ren, S.M. Kachur, R.R. Leadem-Dougherty, S.D. Rhoton, G. Zinser, J. Farlow, P.R. Coker, K.L. Smith, B. Wang, L.J. Kenefic, C.M. Fraser-Liggett, D.M. Wagner, and P. Keim. 2007. Global genetic population structure of *Bacillus anthracis*. *PLoS One* 2(5):e461.

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

van Schaik, W., and R.J. Willems. 2010. Genome-based insights into the evolution of enterococci. *Clin Microbiol Infect* 16(6):527-532.

Varani, A.M., P. Siguier, E. Gourbeyre, V. Charneau, and M. Chandler. 2011. ISsaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. *Genome biology* 12(3):R30.

Werner, G., C. Fleige, U. Geringer, W. van Schaik, I. Klare, and W. Witte. 2011. IS element *IS16* as a molecular screening tool to identify hospital-associated strains of *Enterococcus faecium*. *BMC Infect Dis* 11:80.

Werner, G., R.J. Willems, B. Hildebrandt, I. Klare, andW. Witte. 2003. Influence of transferable genetic determinants on the outcome of typing methods commonly used for *Enterococcus faecium*. *J Clin Microbiol* 41(4):1499-1506.

Willems, R.J., W.P. Hanage, D.E. Bessen, and E.J. Feil. 2011. Population biology of Gram-positive pathogens: high-risk clones for dissemination of antibiotic resistance. *FEMS Microbiol Rev* 35(5):872-900.

Willems, R.J., W. Homan, J. Top, M. van Santen-Verheuvel, D. Tribe, X. Manzioros, C. Gaillard, C.M. Vandenbroucke-Grauls, E.M. Mascini, E. van Kregten, J.D. van Embden, and M.J. Bonten. 2001. Variant *esp* gene as a marker of a distinct genetic

lineage of vancomycin-resistant *Enterococcus faecium* spreading in hospitals. *Lancet* 357(9259):853-855.

Willems, R.J., J. Top, N. van Den Braak, A. van Belkum, H. Endtz, D. Mevius, E. Stobberingh, A. van Den Bogaard, and J.D. van Embden. 2000. Host specificity of vancomycin-resistant *Enterococcus faecium*. *J Infect Dis* 182(3):816-823.

Willems, R.J., J. Top, M. van Santen, D.A. Robinson, T.M. Coque, F. Baquero, H. Grundmann, and M.J. Bonten. 2005. Global spread of vancomycin-resistant *Enterococcus faecium* from distinct nosocomial genetic complex. *Emerg Infect Dis* 11(6):821-828.

Willems, R.J., and W. van Schaik. 2009. Transition of *Enterococcus faecium* from commensal organism to nosocomial pathogen. *Future Microbiol* 4(9):1125-1135.

Williamson, R., C. le Bouguenec, L. Gutmann, and T. Horaud. 1985. One or two low affinity penicillin-binding proteins may be responsible for the range of susceptibility of *Enterococcus faecium* to benzylpenicillin. *J Gen Microbiol* 131(8):1933-1940.

Zapun, A., C. Contreras-Martel, and T. Vernet. 2008. Penicillin-binding proteins and beta-lactam resistance. *FEMS Microbiol Rev* 32(2):361-385.

Zorzi, W., X.Y. Zhou, O. Dardenne, J. Lamotte, D. Raze, J. Pierre, L. Gutmann, and J. Coyette. 1996. Structure of the low-affinity penicillin-binding protein 5 PBP5fm in wild-type and highly penicillin-resistant strains of *Enterococcus faecium*. *J Bacteriol* 178(16):4948-4957.

Zuckerkandl, E., Pauling, L. 1965. Evolutionary divergence and convergence in proteins. Vogel VBaHJ, editor. Academic Press, New York.

VITA

Jessica Galloway-Peña was born in San Antonio, Texas on August 22, 1985, the daughter of Carol Michele Galloway and Josef Peña. After completing the requirements for her high school diploma at Tom C. Clark High School, San Antonio, Texas in May, 2003, she entered Our Lady of the Lake University in San Antonio, Texas, and received a degree of Bachelor of Arts with a major in Biology and a minor in Chemistry from Our Lady of the Lake University in December, 2006. For the next 5 months, she worked as a Special Education Certified Teaching Assistant at Nichols Elementary in San Antonio, Texas. She then took a position as a Research Assistant in May, 2007 in the laboratory of Dr. Steven Norris at The University of Texas Medical School in the Department of Pathology and Laboratory Medicine before entering The University of Texas Health Science Center at Houston Graduate School of Biomedical Sciences in August 2007. In May, 2008 she officially joined the Microbiology and Molecular Genetics Program and chose the laboratory of Dr. Barbara E. Murray in which to do her Ph.D. dissertation research and training.