Microbial Identification and Strain Typing by Transflection Fourier Transform Infrared Spectroscopy with Specific Application to Clinically Relevant Antibiotic Resistant Gram-Positive Bacteria

Ph.D. Thesis

By

Tamao Tsutsumi

Department of Food Science and Agricultural Chemistry

Macdonald Campus, McGill University

Montreal, Quebec

2021

Short Title:

Rapid Microbial Identification and Strain Typing by Transflection FTIR Spectroscopy

Abstract

Microbial identification and strain typing by Fourier transform infrared (FTIR) spectroscopy for clinical diagnostic applications have been demonstrated over the past decade using the transmission and attenuated total reflection modes of spectral acquisition. The transflection (Tr) mode has been less explored for such applications but may offer significant advantages in relation to routine implementation because spectra can be acquired from samples deposited on low-cost disposable slides made from infrared reflective materials such as low-emissivity glass. To evaluate the capabilities of Tr-FTIR spectroscopy for microbial discrimination and identification at the genus and species levels, a spectral database comprised of Tr-FTIR spectra of bacteria and yeasts was constructed, and multivariate analysis methods were applied to these spectral data to develop multi-tier classification models for identification of unknowns. A validation study with 1103 isolates of *Enterococcus faecalis, E. faecium* and *Staphylococcus aureus*, which are leading causes of hospital-acquired infections, yielded over 98% concordance with the results of reference methods. Isolates belonging to other genera or species with less representation in the Tr-FTIR spectral database did not achieve highly concordant results in the validation study. Although increased concordance is expected with increased representation of these species in the spectral database, further investigations is required to determine the extent of the method's discriminatory capabilities in these cases. The Tr-FTIR spectral database was also shown to be capable in identification of isolates cultured on antibiotic-containing media, as part of a novel FTIR-based method for the discrimination of methicillin-resistant *S. aureus* (MRSA) from methicillin-sensitive *S. aureus* and coagulase-negative staphylococcal species. MRSA identification was achieved with 100% sensitivity and 95.3% specificity after 24-h growth of isolates (*n = 56*) on agar containing cefoxitin at a concentration of 4 µg/ml. Strain typing by Tr-FTIR spectroscopy was evaluated using vancomycin-resistant *E. faecium* (VRE) isolates acquired from patient and environmental screening samples and compared against pulsed-field gel electrophoresis (PFGE). Spectral analyses on isolates belonging to two pulsotypes, AA and CC, identified sufficient differences for strain typing, resulting in 91.7% and 91% categorical agreement respectively. The method was also evaluated in a 7-month prospective study, where collected VRE isolates from 2 hospitals were routinely analysed and reported on a weekly basis. At the end of the study, among the 23 clusters identified by Tr-FTIR spectroscopy, 11 (47%), and 6 (26%) clusters yielded 90-100% and 80-90% categorical agreement with PFGE results. Endemic strains that likely resulted in patient colonization were identified based on spectral clustering of environmental samples with those from patients. These studies revealed the usefulness of FTIR spectral profiling as a rapid routine strain typing method for VRE outbreak detection. High-resolution magic-angle spinning nuclear magnetic resonance spectroscopy was used to acquire ^{1}H and ^{31}P NMR spectra from live bacteria as proof of concept for its ability to discriminate at species and subspecies level. Proton and ^{31}P NMR spectra of VRE isolates belonging to pulsotypes AA and CC revealed potential biomarkers related to differences between the pulsotypes. Metabolomic analyses identified increased levels of phospholipids in isolates belonging to the CC pulsotype. Although the identified biomarkers are likely not the sole differences between these VRE strain types, key FTIR spectral regions identified for differentiation between AA and CC pulsotypes correlate with peaks unique to these phospholipids. Overall, the results from this thesis demonstrated the capabilities and limitations of using Tr-FTIR spectroscopy for clinical diagnostic and strain typing applications.

Résumé

L'identification microbienne et le typage des souches par spectroscopie infrarouge à transformée de Fourier (FTIR) pour des applications de diagnostic clinique ont été démontrés au cours de la dernière décennie en utilisant les modes de transmission et de réflexion totale atténuée. Le mode de transflexion (Tr) a été moins exploré pour de telles applications mais peut offrir des avantages significatifs par rapport à la mise en œuvre de routine car les spectres peuvent être acquis à partir d'échantillons déposés sur des lames jetables à faible coût fabriquées à partir de matériaux réfléchissant les infrarouges. Pour évaluer les capacités de la spectroscopie Tr-FTIR pour la discrimination et l'identification microbiennes au niveau du genre et de l'espèce, une base de données spectrale composée de spectres Tr-FTIR de bactéries et de levures a été construite et des méthodes d'analyse multivariée ont été appliquées à ces données spectrales pour développer des modèles de classification à plusieurs niveaux pour l'identification des inconnues. Une étude de validation avec des isolats *d'Enterococcus faecalis*, d'*E. faecium* et de *Staphylococcus aureus* a donné plus de 98 % de concordance avec les résultats des méthodes de référence. Les isolats appartenant à d'autres espèces moins représentés dans la base de données spectrale n'ont pas obtenu de résultats très concordants. Bien qu'une concordance accrue puisse être attendue avec une représentation accrue de ces espèces dans la base de données spectrale, des investigations supplémentaires seraient nécessaires pour déterminer l'étendue des capacités discriminatoires de la méthode dans ces cas. La méthode Tr-FTIR a également correctement identifié les isolats cultivés sur des milieux contenant des antibiotiques, dansd'une nouvelle méthode basée sur le FTIR pour la discrimination de *S. aureus* résistant à la méthicilline (SARM) de *S. aureus* sensible à la méthicilline et des espèces de staphylocoques à coagulase négative. Une sensibilité de 100 % et une spécificité de 95,3 % ont été obtenues pour l'identification du SARM (n = 56) sur gélose avec 4 µg/ml de céfoxitine. Le typage des souches par spectroscopie Tr-FTIR a été évalué à l'aide d'isolats d'*E. faecium* résistants à la vancomycine (ERV) acquis à partir d'échantillons de dépistage de patients et d'environnementaux et comparé à l'électrophorèse sur gel en champ pulsé (PFGE). Les analyses spectrales sur deux pulsotypes, AA et CC, ont identifié des différences pour le typage des souches, résultant en un accord catégorique de 91,7 % et 91 % respectivement. La méthode a été évaluée prospectivement pour 7 mois; les ERV collectés ont été analysés et rapportés sur une base hebdomadaire, et comparés rétrospectivement aux résultats de PFGE. Parmi les 23 clusters identifiés par spectroscopie Tr-FTIR, 11 (47 %), et 6 (26 %) clusters ont donné un accord catégorique de 90 à 100 % et 80 à 90 % avec les résultats de la PFGE. Ces études ont révélé l'utilité du spectroscopie FTIR en tant que méthode rapide de typage des souches de routine pour la surveillance. Les spectres RMN à haute résolution de rotation à l'angle magique des bactériennes vivantes ont démontré les capacités à discriminer au niveau des espèces et des sous-espèces. Les spectres RMN d'isolats ERV appartenant aux pulsotypes AA et CC ont révélé des biomarqueurs potentiels liés aux différences dans ces pulsotypes. Les analyses métabolomiques ont identifié une présence accrue des phospholipides dans le CC. Bien que ces biomarqueurs identifiés ne soient probablement pas les seules différences entre ces isolats, leur contribution à la différenciation entre les pulsotypes AA et CC par spectroscopie FTIR est cohérente avec les régions spectrales FTIR jugées essentielles pour une différenciation réussie. Les résultats de cette thèse ont démontré les capacités et les limites de l'utilisation de la spectroscopie Tr-FTIR pour les applications de diagnostic microbien clinique et de typage des souches.

Acknowledgements

First and foremost, I would like to thank Dr. Ismail for the patience and guidance over the years throughout my undergraduate and graduate studies. Without his encouragement to pursue research in his laboratory some years ago, I would not have been able to experience the many opportunities he has provided me.

Special thank you to Dr. Pierre Lebel for providing initial access for the research at McGill University Health Centre (MUHC) through collaboration with Dr. Ismail. To Dr. Charles Frenette, the Medical Director of Infection Control at MUHC, who provided me with guidance on the importance and the complexity of conducting surveillance for infectious disease related pathogens in hospital settings. To Mrs. Nancy Doherty, who taught me the technical aspects of infection control and surveillance, from data acquisition in the microbiology laboratory to data analysis. Mr. Martin Raymond and all the laboratory technicians at the MUHC microbiology laboratory for assisting me in the training and acquisition of data of bacterial isolates.

I would like to thank you the many research collaborators who welcomed me into their laboratories for access to their bacteria samples. This includes Dr. Émilie Vallières at Centre hospitalier universitaire – Sainte-Justine (Montreal, Canada), Dr. George Golding at the National Microbiology Laboratory (Winnipeg, Canada), Mrs. Irene Iugovas at Health Canada (Longueil, Canada), Dr. Catherine Carrillo at Canadian Food Inspection Agency (Ottawa, Canada) and Dr. Michelle Hill at QIMR Berghofer Medical Research Institute (Brisbane, Australia).

Thank you to the Drug Discovery Platform at MUHC-Research Institute for their services in providing me access to NMR spectroscopy and imaging mass spectrometry for biomarker analysis. Dr. Andrée Gravel for teaching me how to acquire and analyze data acquired by NMR spectroscopy, and Dr. Nidia Lauzon for assisting with the imaging mass spectrometry data acquisition and analysis.

Thank you to Dr. Catherine Carrillo at Canadian Food Inspection Agency (Ottawa, Canada), for not only providing bacteria isolates for FTIR spectral acquisition, but also educating and guiding me on whole genome sequencing with her knowledge, insight.

Thank you to Dr. Jacqueline Sedman for the insightful scientific guidance, in particular on the overall thesis direction, as well as knowledge on spectroscopy and chemometrics. A very

special thank you to my colleague Dr. Lisa Lam, for always being very supportive and a source of motivation for me throughout the years. The countless hours spent in the lab together, all of the scientific discussion, and adventures at numerous conferences over the years were truly an irreplaceable and important part of my thesis experience. My Ph.D. experience would not have nearly been as interesting without you. I would also like to thank the rest of my current and former colleagues, Dr. Mazen Bahadi, Dr. Olivia Tao, Hayline Kim, Xin Di Zhu and Ivan Travcar, for all the advice, conversations, and technical help in the lab. Thank you to my family who have supported me throughout the years with an incredible amount of encouragement and patience.

Finally, I would like to thank the Natural Sciences and Engineering Research Council of Canada for the financial support throughout my graduate studies with the CGS-M and PGS-D scholarships, and McGill University for the Graduate Mobility Award, allowing me to fully focus on my research and academic progress.

Contributions to Knowledge

This thesis introduced novel methods utilizing transflection-FTIR spectroscopy and HR-MAS NMR spectroscopy for microbial identification and strain typing. These include the development of a spectral database for species identification, a strain typing method for outbreak detection, as well as a new method to accurately identify MRSA using FTIR spectroscopy. The primary contributions to knowledge resulting from this research are summarized below:

1. **Comparison of transflection FTIR spectral acquisition method against ATR- and transmission modes in microbial discrimination.**

   The three spectral acquisition modes were compared for their ease of sample preparation, data acquisition and data analyses, as well as spectral quality for application in microbial discrimination, using a common set of microorganisms. Comparable spectral quality and discriminatory capabilities were observed in the transflection mode, relative to ATR and transmission spectral acquisition modes.

2. **Development and evaluation of a transflection-FTIR spectral database encompassing Gram-positive and Gram-negative bacteria and yeasts for reliable identification of certain microorganisms.**

   Transflection FTIR spectra of clinically relevant isolates, with a focus on Gram-positive species, were collected and a classification system for microorganisms based on spectral differences was developed. Spectral data acquired at local and international laboratories showcased that with standardized sample preparation and sufficient species representation in the spectral database, correct identification to the species level was possible, regardless of the specimen source such as clinical, food or veterinary origin.

3. **Optimization of MRSA detection using FTIR spectroscopy through the addition of antibiotics to culture media.**

   The method for MRSA identification using FTIR spectroscopy was developed by combining the capability of FTIR-spectroscopy to discriminate *S. aureus* from CoNS species, and the ability to selectively grow methicillin-resistant staphylococci on a culture medium that contains the antibiotic cefoxitin. A bi-plate concept, where staphylococcal isolates were cultured on blood agar plates with and without cefoxitin, allowed for detection of both sensitive and resistant strains. All samples with positive growth were subjected to spectral analyses, for identification of *S. aureus.*

4. **Development and evaluation of a routine strain typing method using transflection-FTIR spectroscopy to detect nosocomial outbreaks, and to continuously track the transmission of pathogens.**

   Transflection-FTIR spectra acquired from VRE isolates with previously identified PFGE pulsotypes were analysed using supervised and unsupervised multivariate statistical analysis techniques. Successful differentiation of VRE isolates based on pulsotypes as well as retrospective identification of clonal isolates was achieved. A method for identifying spectrally indistinguishable isolates based on the degree of similarity of FTIR spectral characteristics was developed, and was tested in a prospective 6-month study, for two Montreal hospitals; the method showcased the benefit of routinely analyzing data, rather than in retrospect for VRE outbreak detection.

5. **Demonstration of microbial discrimination using spectra acquired by HR-MAS NMR spectroscopy**

   Bacterial colonies were directly analyzed using ^{1}H and ^{31}P HR-MAS NMR spectroscopy, as a proof of concept, showcasing the capability to discriminate between species and strain types, similarly to FTIR spectroscopy. As demonstrated with transflection FTIR spectroscopy, spectral acquisition and analysis conducted on live cells by HR-MAS NMR spectroscopy has the potential to become more relevant as a diagnostic tool in clinical microbiology upon reductions in instrument size, and cost and in sample size, together with automation and standardization of the method. Differences observed in the ^{1}H and ^{31}P HR-MAS NMR spectra between VRE pulsotypes tentatively identified biomolecules that may have contributed to FTIR spectral differences between these pulsotypes.

6. **Preliminary data acquisition using high resolution magic angle spinning nuclear magnetic resonance (HR-MAS NMR) spectroscopy, mass spectrometry (MS) and whole genome sequencing for biomarker elucidation and correlation to FTIR spectral strain typing data**

   Based on the FTIR and HR-MAS NMR spectral analyses of VRE isolates, preliminary analysis on biomarker elucidation was conducted to identify biomolecules contributing to successful spectral discrimination between VRE isolates of different pulsotypes. Whole genome sequencing data, as well as MS and NMR spectra of extracted chloroform-soluble fractions were obtained to gain insight into the genotypic and biomolecular differences. Phosphorus-

containing biomarker(s) as well as choline-containing components were tentatively identified as potential biomarkers in discriminating between the studied VRE strain types. The genotypic and spectroscopic data acquired contributed toward understanding the biochemical differences that were observed in FTIR spectral data for successful strain type discrimination.

Contribution of Authors

Chapters 3-6 of this thesis are texts of papers being prepared for publication, and have been presented at conferences as oral and/or poster presentations. The author was responsible for the concepts, design of experiments, experimental work, and manuscript preparation. Dr. Ashraf A. Ismail is the thesis supervisor, and had direct advisory input into the work. Dr. Jacqueline Sedman provided valuable advice and insight to all aspect of the thesis, particularly on spectroscopy. Dr. Lisa Lam, Xin Di Zhu, Liam Liang, Dr. Alok Shah, Stefani Mazurkiewicz, assisted in FTIR spectral data acquisition. Dr. Émilie Vallières at Centre hospitalier universitaire – Sainte-Justine (Montreal, Canada), Dr. George Golding at the National Microbiology Laboratory (Winnipeg, Canada), Mrs. Irene Iugovas at Health Canada (Longeuil, Canada), Dr. Catherine Carrillo at Canadian Food Inspection Agency (Ottawa, Canada), Dr. Michelle Hill at QIMR Berghofer Medical Research Institute (Brisbane, Australia) for providing well-characterized isolates for this research, and the microbiological data presented in Chapters 3. Mr. Martin Raymond provided isolates and microbiological data presented in Chapter 4. Mrs. Nancy Doherty provided vancomycin resistant *Enterococcus faecium* isolates and their genotypic strain typing data and epidemiological data as part of research described in Chapter 5. Dr. Catherine Carrillo, Dr. Andrée Gravel, Dr. Nidia Lauzon, Mrs. Nancy Doherty acquired data by whole genome sequencing, nuclear magnetic resonance spectroscopy, matrix assisted laser desorption/ionization – time of flight mass spectrometry imaging, and pulsed gel electrophoresis, respectively, as part of the research described in Chapter 6.

Chapter 3.

Poster Presentation

Tamao Tsutsumi, Alok K. Shah, Lisa Lam, Sanmarie Schlebusch, Annika Krueger, Ian H. Frazer, Peter Hugenholtz, HP Soyer, M Morrison, Michelle M. Hill, Jacqueline Sedman, Ashraf A. Ismail.

Staphylococcus Species Identification by Fourier Transform Infrared (FTIR) Spectroscopic Techniques: A Cross-Lab Study (IDWeek2019, Washington D.C., USA October 1-7, 2019)

Chapter 4.

Poster Presentation

Tamao Tsutsumi, Charles Frenette, Nancy Doherty, Jacqueline Sedman, Ashraf A. Ismail.

Rapid identification of methicillin-resistant Staphylococcus aureus (MRSA) in clinical microbiology labs by infrared spectral fingerprinting following growth on agar supplemented with cefoxitin. (ECCMID 2020 - cancelled, Paris, France April 18-21, 2020)

Chapter 5.

Tamao Tsutsumi, Charles Frenette, Nancy Doherty, Jacqueline Sedman, Ashraf A. Ismail.

Rapid and Reagent-Free Typing of Vancomycin-Resistant Enterococcus faecium by Transflection-Fourier Transform Infrared (FTIR) Spectroscopy as a New Tool for Outbreak Investigations

Poster Presentation

Tamao Tsutsumi, Charles Frenette, Nancy Doherty, Jacqueline Sedman, Ashraf A. Ismail.

Transflection Fourier Transform Infrared Spectroscopy as a Real-Time Strain Typing Technique: A Vancomycin Resistant Enterococcus faecium (VRE) Typing Prospective Study (IDWeek2019, Washington D.C., USA October 1-7, 2019)

Chapter 6.

Poster Presentations

Tamao Tsutsumi, Charles Frenette, Andree Gravel, Nancy Doherty, Jacqueline Sedman, Ashraf A. Ismail.

Rapid and Reagent-Free Typing of Vancomycin-Resistant Enterococcus faecium by Whole Organism Fingerprinting Techniques: New Tools for Outbreak Investigations (MOOT31 NMR symposium Sherbrooke, QC, October 19-21 2018)

Rapid Strain Typing by High-Resolution Magic Angle Spinning Nuclear Magnetic Resonance Spectroscopy (American Society of Microbiology (ASM) 2019, San Francisco, USA, June 20-25, 2019)

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Full Name
ANN	artificial neural network
ASM	American Society of Microbiology
AST	antimicrobial susceptibility test
ATCC	American Type Culture Collection
ATR	attenuated total reflectance
BAP	blood agar plate
BAPO	Blood agar plate, Oxoid
BAPT	Blood agar plate, Tamao
BHI	Brain Heart Infusion
CA	community acquired
CaF_2	calcium fluoride
CDC	Center for Disease Control and Prevention
CFIA	Canadian Food Inspection Agency
cgMLST	core genome multilocus sequence typing
cgSNP	core genome single nucleotide polymorphism
CHUSJ	Centre hospitalier universitaire Sainte-Justine
CL	cardiolipin
CLSI	Clinical and Laboratory Standards Institute
CMRSA	Canadian methicillin resistant *Staphylococcus aureus*
CNA	colistin-nalidixic acid
CNAO	Blood agar plate with colistin, nalidixic acid, Oxoid
CNAT	Blood agar plate with colistin, nalidixic acid, Tamao
CONS	coagulase negative *Staphylococci*
CPA	cyclopropanic acid
CRE	carbapenem-resistant Enterobacteriaceae
CRISPR	clustered regularly interspaced short palindromic repeats
DAN	1,5-diaminonaphtalene
DHB	2,5-dihydroxybenzoic acid

DNA	deoxyribonucleic acid
DTGS	deuterated triglycine sulfate
EDTA	ethylene diamine tetra acetic acid
ESBL	extended-spectrum beta-lactamase
EUCAST	European Committee on Antimicrobial Susceptibility Testing
FOX	cefoxitin
FPA	focal plane array
FTIR	Fourier transform infrared
GN	Gram-negative
GP	Gram-positive
GP-DGDAG	glycerophosphoryldiglucosyldiacylglycerol
HAI	Hospital acquired infection
HCA	hierarchical cluster analysis
HCN	Hydrogen-Carbon-Nitrogen
HCP	Hydrogen-Carbon-Phosphorous
HR-MAS NMR	high resolution magic angle spinning nuclear magnetic resonance
ICU	Intensive Care Unit
ID	identification
IMS	imaging mass spectrometry
IR	infrared
ISO	International Organization for Standardization
KB	kilobyte
KNN	K-nearest neighbour
KPC	*Klebsiella pneumoniae* carbapenemase
LA	livestock-associated
LA-FTIR	linear array-Fourier transform infrared
LC	liquid chromatography
low-E glass	low-emissivity glass
LSPQ	Laboratoire de santé publique du Québec
MALDI-TOF MS	matrix assisted laser desorption/ionization time of flight mass spectrometry
MB	megabyte

Mbp	mega base pairs
MCT	mercury cadmium telluride
MGH	Montreal General Hospital
MIC	minimum inhibitory concentration
MLST	multilocus sequence typing
MLVA	multi-locus variable number of tandem repeats analysis
MRS	methicillin resistant staphylococci
MRSA	methicillin resistant *Staphylococcus aureus*
MSCRAMM	microbial surface components recognizing adhesive matrix molecules
MSS	methicillin sensitive staphylococci
MSSA	methicillin sensitive *Staphylococcus aureus*
MUHC	McGill University Health Centre
NGS	next generation sequencing
NIR	near infrared
NML	National Microbiology Laboratory
PBP	penicillin-binding protein
PC	principle component
PC	phosphatidylcholine
PCA	principle component analysis
PCR	Polymerase chain reaction
PE	phosphatidylethanolamine
PFGE	pulsed field gel electrophoresis
PG	phosphatidylglycerol
PI	phosphatidylinositol
PLS-DA	partial least-squares-discriminant analysis
PSM	phenol soluble modulin
PVL	Panton-Valentine leucocidin
QIMRB	Queensland Institute of Medical Research Berghofer
RMS	root mean square
rRNA	ribosomal ribonucleic acid
RVH	Royal Victoria Hospital

SCCmec	staphylococcal cassette chromosome mec
SCV	small colony variant
SED	single element detector
SMRT	single-molecule real time
SNR	Signal-to-noise ratio
SNV	single nucleotide variant
SOM	self-organizing map
SVM	support vector machine
TFA	trifluoroacetic acid
TPP	triphenyl phosphate
TSP	3-(Trimethylsilyl)propanoic acid
VISA	vancomycin intermedia *Staphylococcus aureus*
VRE	vancomycin resistant enterococci
VSE	vancomycin sensitive Enterococci
wgMLST	whole genome multilocus sequence typing
WGS	whole genome sequencing
WHO	World Health Organization
ZnSe	Zinc selenide

Chapter 1. Introduction

1.1 General Introduction

Hospital associated outbreaks and hospital acquired infections are critical problems, and have serious consequences to patients staying at hospital, increasing their mortality and morbidity rates [1]. Pathogens causing nosocomial infections and outbreaks are increasingly becoming antibiotic resistant as a result of increased exposure to antibiotics within hospitals [2]. It is critical to actively screen patients at the time of admission and periodically throughout their hospitalization to monitor the presence of antibiotic resistant pathogens such as methicillin-resistant *Staphylococcus aureus* (MRSA) and vancomycin-resistant *Enterococcus faecium* (VRE) as part of infection control and nosocomial outbreak surveillance. In Canada and many countries around the world, VRE is one of the Gram-positive pathogens that is commonly isolated from the hospital environment, and is often considered as an indicator for insufficient adherence to sanitation protocols [3]. VRE screening samples must be identified using a combination of techniques, such as screening broth, PCR for vancomycin resistance gene detection, and isolation on chromogenic agar, which can take up to three days cumulatively. Upon confirmation of VRE presence relatedness between isolates is investigated via strain typing. Pulsed field gel electrophoresis (PFGE), the current gold standard for strain typing, is a time-consuming and laborious technique, and the slow turnover rate for results by PFGE hinders the infection control team to implement actions necessary based on results from both molecular and epidemiological information. PFGE remained to be the gold standard for over two decades, despite its limited resolution, slow-turnaround time to results, and lack of reproducibility related to the degree of difficulty in the method. Although whole genome sequencing (WGS), another genotypic method, is starting to replace PFGE as the standard for strain typing, the lack of standardized and streamlined processes and the burden of computing enormous amounts of data are currently hindering its routine use [4]. In order to better track the presence and transmission of pathogenic organisms within hospitals, rapid microbial identification and routine strain typing are necessary.

Whole-organism fingerprinting techniques, such as Fourier transform infrared (FTIR) spectroscopy and nuclear magnetic resonance (NMR) spectroscopy, acquire spectral data of microorganisms with minimal to no sample preparation. Both FTIR and NMR spectra acquired in this manner characterize the biochemical makeup of the intact microbial cells, and the sensitivity of these whole-organism fingerprinting techniques to biochemical differences between

microorganisms as been shown to be sufficient for subspecies level discrimination of microorganisms [5-10]. An important benefit of these techniques is that the spectra acquired from samples can be used in many applications, such as identification and strain typing, through the use of various multivariate statistical analyses. FTIR spectroscopy in particular, has the potential to be implemented as a rapid screening method, with its fast data acquisition capabilities.

Transflection is a spectral acquisition mode for FTIR spectroscopy that has been less commonly studied for microbial applications in comparison to transmission and attenuated total reflectance (ATR) acquisition modes. For spectral acquisitiom in the transflection mode, the sample is placed on an infrared-reflective substrate, and the infrared bean is transmitted through the sample, reflected back from the surface of the substrate, and transmitted through the sample for the second time and then directed by the optics to the infrared detector. The transflection mode has been less popular due to reported spectral distortion caused by the differences in the refractive index changes at the air/sample and sample/substrate interfaces [11]. However, there are also reports indicating that mathematical processing of the spectra prior to analyses can eliminate such distortions, and enable successful qualitative analyses [12, 13]. The substrate on which samples are deposited for transflection mode spectral acquisition is a cheaper alternative compared to the infrared-transparent windows used for to acquire transmission spectra, thereby making transflection a more cost-effective method compared to transmission mode [12]. Furthermore, the cheaper costs enable the slides to be used as a one-time-disposable consumable, making handling of pathogenic microbial samples for lab personnel, while eliminating the protocols required for reusing a more expensive substrate. In addition, the transflection mode has the potential for automated spectral acquisition with a motorized stage, which is not as easy to implement for the ATR mode. In this thesis, the main focus was to evaluate transflection FTIR spectroscopy for microbial identification and strain typing applications, with a focus on the most prevalent antibiotic-resistant Gram-positive pathogens, namely VRE and MRSA.

High-resolution magic-angle spinning (HR-MAS) NMR spectroscopy is a type of NMR spectral acquisition technique that is employed to acquire highly resolved NMR spectra from live microbial cells, enabling direct analysis of microbial samples in their native state. Applications of HR-MAS NMR spectroscopy for characterization of cell-surface polysaccharides and strain types from whole cells [14]. NMR spectroscopy is commonly used in metabolomics in combination with mass spectrometry [15]. Like FTIR spectral data, spectra acquired by HR-MAS NMR

spectroscopy provide a snapshot on the metabolic and structural state of the live microbial cells. By taking advantage of the higher chemical specificity that NMR spectroscopy offers, identification of specific biomarkers may be achieved in metabolomic studies by HR-MAS NMR spectroscopy may be correlated to FTIR spectral data.

1.2 Overview and Objectives of the Research

1.2.1 Goal of the Research

The main objective of this thesis was to evaluate transflection FTIR spectroscopy, as a rapid technique for identification and strain typing of clinically relevant Gram-positive bacteria. The research included development and evaluation of a transflection FTIR spectral database for species identification and a FTIR spectroscopy based strain typing method. A key objective was to showcase the benefits of using spectroscopic techniques as rapid, screening tools, at a time where genotypic techniques are gaining popularity. The focuses were to a) develop a transflection FTIR-based spectroscopic technique that can be integrated into the current workflow in microbial diagnostic laboratories for both identification and strain typing, b) to optimize the culturing conditions of bacteria to improve FTIR spectral identification and strain typing of antibiotic-resistant microorganisms, with MRSA as an example, and c) to demonstrate and use HR-MAS NMR spectroscopy for studying live bacteria cells for subspecies spectral discrimination, as well as biomarker elucidation in combination with other spectroscopic and genotypic methods. The FTIR spectra of microorganisms represent their biochemical status, and therefore in combination with various multivariate statistical analysis tools, the spectra can provide useful data for characterization at multiple taxonomic levels.

1.2.2 Specific Objectives of the Research

(i) To develop and evaluate a transflection-FTIR spectral database for microbial identification with isolates obtained from clinical and food sources

(ii) To develop and evaluate a method for rapid and accurate identification of MRSA using FTIR spectroscopy in combination with selective media, as a cheaper alternative method to the use of chromogenic agar

(iii) To develop and evaluate a protocol for determining strain relatedness between nosocomial VRE strains using FTIR spectroscopy, for routine outbreak detection and surveillance applications

(iv) To demonstrate the potential of ^{1}H and ^{31}P HR-MAS NMR spectroscopy in microbial characterization, discrimination and strain typing from live microbial cells

(v) To attempt correlating between NMR and FTIR spectral data to elucidate biomarkers associated within FTIR spectral discrimination between VRE isolates at the strain level, in combination with mass spectrometry and whole genome sequencing

1.3 References

1. Organization, W.H., *Antimicrobial resistance: global report on surveillance*. 2014: World Health Organization.
2. CDC, *Antibiotic resistance threats in the United States, 2019*. 2019, U.S. Department of Health and Human Services: Atlanta, GA.
3. Zhou, X., et al., *Enterococcus faecium: from microbiological insights to practical recommendations for infection control and diagnostics.* Antimicrobial Resistance & Infection Control, 2020. **9**(1): p. 1-13.
4. Freitas, A.R., et al., *Multidrug-resistant high-risk Enterococcus faecium clones: can we really define them?* International Journal of Antimicrobial Agents, 2020: p. 106227.
5. Amiali, N.M., et al., *Evaluation of Fourier transform infrared spectroscopy for the rapid identification of glycopeptide-intermediate Staphylococcus aureus.* Journal of antimicrobial chemotherapy, 2007. **61**(1): p. 95-102.
6. Amiali, N.M., et al., *Epidemiological typing of methicillin-resistant Staphylococcus aureus strains by Fourier transform infrared spectroscopy.* Journal of microbiological methods, 2007. **69**(1): p. 146-153.
7. Kirkwood, J., et al., *Differentiation of group I and group II strains of Clostridium botulinum by focal plane array Fourier transform infrared spectroscopy.* Journal of food protection, 2006. **69**(10): p. 2377-2383.
8. Vogt, S., et al., *Fourier-transform infrared (FTIR) spectroscopy for typing of clinical Enterobacter cloacae complex isolates.* Frontiers in microbiology, 2019. **10**.
9. Rodrigues, C., et al., *On the front line of Klebsiella pneumoniae surface structures understanding: establishment of Fourier Transform Infrared (FT-IR) spectroscopy as a capsule typing method.* bioRxiv, 2019: p. 554816.
10. Gudlavalleti, S.K., et al., *In vivo determination of Neisseria meningitidis serogroup A capsular polysaccharide by whole cell high-resolution magic angle spinning NMR spectroscopy.* Carbohydrate research, 2006. **341**(4): p. 557-562.
11. DeVetter, B.M., et al., *Characterization of the structure of low-e substrates and consequences for IR transflection measurements.* Vibrational Spectroscopy, 2017. **91**: p. 119-127.
12. Baker, M.J., et al., *Using Fourier transform IR spectroscopy to analyze biological materials.* Nature protocols, 2014. **9**(8): p. 1771.
13. Wehbe, K., et al., *The effect of optical substrates on micro-FTIR analysis of single mammalian cells.* Analytical and bioanalytical chemistry, 2013. **405**(4): p. 1311-1324.
14. Li, W., et al., *Methods for acquisition and assignment of multidimensional high-resolution magic angle spinning NMR of whole cell bacteria.* Analytical chemistry, 2005. **77**(18): p. 5785-5792.
15. Bingol, K. and R. Brüschweiler, *Two elephants in the room: new hybrid nuclear magnetic resonance and mass spectrometry approaches for metabolomics.* Current opinion in clinical nutrition and metabolic care, 2015. **18**(5): p. 471.

Chapter 2. Literature Review

2.1 Abstract

Nosocomial outbreaks and hospital acquired infections are critical problems, that have serious consequences to patients staying at hospitals or long term acute care facility. Pathogens causing nosocomial infections and outbreaks are increasingly becoming antibiotic resistant as a result of increased exposure to antibiotics within the hospitals. It is critical to screen patients at the time of admission and continue to do so periodically throughout their hospitalization to monitor the presence of microorganisms such as methicillin resistant *Staphylococcus aureus* and vancomycin resistant enterococci as part of infection control and nosocomial outbreak surveillance. Combinations of techniques are used to identify and strain type the target organisms. Pulsed field gel electrophoresis (PFGE), the current gold standard for strain typing is a time-consuming and laborious technique. The slow turnover rate for results by PFGE hinders the infection control team to implement optimal necessary actions, as it is only able to provide results retrospectively. Although whole genome sequencing (WGS) is starting to replace PFGE as the standard for strain typing, the lack of streamlined processes and the burden of computing enormous amounts of data is currently hindering its use routinely. In order to minimize transmission of pathogenic organisms between patients in hospitals, rapid microbial identification and strain typing is necessary. Whole-organism fingerprinting techniques, such as Fourier transform infrared (FTIR) spectroscopy and nuclear magnetic resonance (NMR) spectroscopy acquire data based on the biochemical makeup of live organisms with minimal to no sample preparation once colonies are isolated. FTIR and NMR spectroscopy both have been shown to have sufficient sensitivity for subspecies level discrimination. Furthermore, by exploiting the abundant spectral information, each acquired spectrum can be used for identification and strain type characterization with use of multivariate statistical analysis methods, thereby reducing the time and expertise required to conduct multiple techniques for microbial characterisation. The gap in routine strain typing could potentially be filled by using FTIR spectroscopy. Furthermore, by utilizing complementary techniques such as NMR spectroscopy, in combination with genotypic and other spectroscopic techniques, further understanding on the biochemical differences attributed to metabolic or structural differences and changes in live microbial cells that are reflected in the FTIR spectra can be correlated.

2.2 Introduction

In healthcare settings, the identification of bacteria is necessary to enable doctors to appropriately treat infections, or to track hospital acquired colonization and infection, to implement infection prevention and control protocols to minimize or intercept transmission of pathogens within healthcare facilities. Antibiotic resistant bacteria are becoming more ubiquitous, found in animals (both domesticated and wild), humans, and in food [1-6]. Pathogens associated with hospital acquired infections are often antibiotic resistant, such as vancomycin resistant *Enterococcus faecium* (VRE), methicillin resistant *S. aureus* (MRSA), and extended spectrum β-lactamase (ESBL) producing and carbapenem-resistant Enterobacteriaceae species (CRE). Hospitals create a unique environment for bacteria to gain resistance against multiple antibiotics, caused by frequent exposures to antimicrobial agents [7]. Infection control is crucial in order to track ongoing outbreaks and to prevent new ones from occurring. Nosocomial outbreaks caused by antibiotic resistant pathogens are of particular interest because patients who are affected tend to be at high risk of infection through acquired colonization; these people include those with compromised immune systems, who have undergone surgeries, and of extreme ages (very young or old). It is crucial to identify reservoirs/sources and prevent/stop transmission of nosocomial pathogens within healthcare settings to new patients through the use of active screening as part of infection prevention and control protocols. Surveillances for such pathogens are done at different scales ranging from local (i.e. within the hospital), regional, provincial, national and international levels. Surveillance allows monitoring and detection of outbreaks, based on information obtained from the molecular tests, patient records and space-time epidemiology. Routine screening for universal and targeted admission have been correlated to reduction in prevalence rates for both MRSA and VRE [8]. Molecular typing techniques play an important role in surveillance and outbreak investigation by providing genotypic and/or phenotypic information regarding the pathogen, enabling comparison between isolates obtained from patient screening and infection sites. Combinations of techniques are used to identify and confirm its strain type, such as polymerase chain reaction (PCR), selective and differential media, mass spectrometry and pulsed field gel-electrophoresis. These techniques cumulatively are very time consuming and labor intensive. The strain type results is used in retrospect, while infection control protocols are implemented upon presumption that there is an outbreak, when new cases arise from screening at higher than usual (baseline) rate. While being proactive and implementing strategies to prevent

transmission of pathogens within the hospital is important, conducting these procedures when unnecessary (e.g. if there is actually no outbreak going on), can be costly for the hospitals. Actions taken for infection prevention and control include increased frequency for active patient and environmental screening, isolating colonized patients in different rooms, and conducting deep cleaning, to remove pathogens from high-contact surfaces, equipment and linens. If these molecular information on the pathogens can be obtained with a shorter turnaround time, it can be used to optimize when to implement such infection control protocols, and overall, will allow hospitals to track in real time of nosocomial pathogens and outbreaks.

Thus, rapid and accurate identification and strain typing techniques are required to effectively monitor the state of nosocomial pathogens and outbreaks. Whole-organism fingerprinting techniques such as Fourier transform infrared (FTIR) spectroscopy and high-resolution magic angle spinning nuclear magnetic resonance (HR-MAS NMR) spectroscopy are rapid and reagent free techniques that can acquire information from whole, live organisms including bacteria. Spectra acquired from these techniques are spectroscopic fingerprints of the isolates based on their unique biochemical makeup. Among these techniques, FTIR spectroscopy has been the most widely studied, with reports on bacteria identification and sub-species level differentiation dating back to 1950s. Bacterial identification and strain typing by these techniques will be reviewed in context with current techniques, evaluating the potential implementation of FTIR and HR-MAS NMR spectroscopy for routine rapid bacteria identification and strain typing in surveillance agencies and hospitals.

2.3 Antimicrobial-resistant Gram-positive bacteria: MRSA and VRE

MRSA and VRE are the two of the most common nosocomial pathogens in Canada along with *Clostridium difficile* [8]. Median prevalence of MRSA colonization and infection, and VRE colonization and infection per 100 in-patients were 4.1 and 0.8 respectively [8]. MRSA is the leading cause of hospital acquired infection (HAI) in the USA with ~80,500 severe MRSA infections per year reported in 2013 [7]. In Europe, *S. aureus* and enterococcal species are among the top 10 organisms that cause HAI in acute care hospitals [9].

2.3.1 Methicillin-resistant *Staphylococcus aureus* (MRSA)

S. aureus is a commensal Gram-positive bacterium, that colonizes humans in sites such as, but not limited to, the nasal carriage, skin, hair, nail, and rectum [10]. Unlike most other

staphylococcal species, *S. aureus* are known to be invasive, causing infections and toxin-mediated diseases [11]. Invasive *S. aureus* infection is a major cause of morbidity and mortality. *S. aureus* infections can be acquired at hospitals through transmission between patients and/or healthcare personnel. More complicated infections occur when surgical wounds or entrance sites of indwelling intravenous devices become infected. Such infections occur when proper sterilization was not achieved. *S. aureus* is the most common cause of post-operative wound infections worldwide [12]. These infections are difficult to treat, even when the bacteria are antibiotic susceptible, as these sites can become persistently infected. Bloodstream infections are especially very serious as infections can spread throughout the body. Skin and soft tissue infections caused by *S. aureus* are common and tend to result from community-acquired (CA) - MRSA infections, rather than hospital-acquired (HA) – MRSA infections. *S. aureus* produces many types of extracellular toxins, that cause toxin-mediated diseases such as food poisoning, toxic shock syndrome, and staphylococcal scalded skin syndrome [11]. Without appropriate treatment, *S. aureus* infections can result in death.

MRSA is defined as "isolates that carry the *mecA* gene or a related variant known as *mecC*, that confer resistance to all β-lactam antibiotics, including cephalosporins and carbapenems" [13]. Methicillin is a penicillinase-stable antibiotic that was first introduced to treat *S. aureus* infections, but resistance against this antibiotic quickly emerged [14]. Methicillin and other β-lactam antibiotics inhibit cell-wall synthesis by preventing the transpeptidation between peptides of adjacent peptidoglycan strands in *S. aureus*, by acting as a substrate analog of D-Ala-D-Ala peptidoglycan side chain. MRSA is known to be heterogeneously resistant to β-lactams, where most cells display low levels of resistance, with only 0.01-0.1% of the colony displaying high resistance levels. MRSA that express homogenous resistance is rare. MRSA can be induced to express homogeneous resistance through culturing with exposures to β-lactams, and this is reported to be the result of gene expression alteration rather than adaptation to the antibiotics [15]. Cefoxitin, a second-generation cephamycin antibiotic, is commonly used to detect MRSA, and is known to be a better surrogate for identifying MRSA compared to oxacillin, as it improved detection for heterogeneously resistant strains [16]. Growth in the presence of cefoxitin enables for accurate MRSA identification, as MSSA have minimum inhibition concentration (MIC) $\leq$ 4mg/L, whereas MRSA exhibit MIC $\geq$ 8mg/L, most of which have MIC $\geq$ 125mg/L [17].

2.3.1.1 Methicillin resistance mechanism in MRSA

Staphylococcal cassette chromosome mec (*SCCmec*) is a mobile gene element that is only present in *Staphylcoccus* species that are resistant against methicillin. *mecA*, a part of the *mec* gene complex of *SCCmec* encodes the production of penicillin-binding protein 2a (PBP2a), a modified version of the original penicillin-biding protein (PBP), used for transglycosylation and transpeptidation for cell wall synthesis. PBP2a reduces the affinity of the cell wall to almost all β-lactam antibiotics. By doing so, the transpeptidation step in the cell wall synthesis is maintained [15, 18]. The *mecC* gene, like the *mecA* gene is part of *SCCmec*. *mecC* is a relatively new resistance gene discovered in MRSA in 2011. Since then, *mecC* MRSA has been identified in river water [19], and in domesticated animals such deer, cows, and sheep [1, 4, 5]. The *mecC* gene shares 70% of its gene sequence with *mecA*, and the PBP produced by *mecC* is ~ 63% similar to PBP2a. Due to its low genetic similarity to *mecA* gene, *mecC* MRSA cannot be detected using identification techniques developed for the *mecA* gene [14, 20, 21]. To further complicate matters, *mecC* MRSA isolates have low resistance against oxacillin, and therefore can go undetected during antibiotic susceptibility testing. Dupieux et al. (2017) tested the performance of four different brands of antibiotic containing chromogenic agar for *mecC* MRSA detection, and reported that depending on the brand of the chromogenic agar, the sensitivity level for *mecC* MRSA varied from 63% to 99% [20].

2.3.1.2 MRSA strain types

MRSA is generally categorized into three groups, hospital-associated, community-associated and livestock-associated, depending on where the MRSA originates or its reservoir. *S. aureus* evolved and adapted to various environmental conditions, which became separate lineages, through acquisition of *SCCmec* and other mobile genetic elements. CA-MRSA was discovered when a patient without a record of hospital admission was diagnosed with MRSA, in the 1990s. Since then, CA-MRSA has been categorized separately from HA-MRSA, as it had different antibiotic susceptibility profiles, and virulence. CA-MRSA can infect the people who are young, and healthy, with low risks of acquiring HA-MRSA infections, while HA-MRSA cause invasive infections in immunocompromised patients, during or shortly after their visit or stay at healthcare facilities [22, 23]. CA-MRSA is often associated with skin and soft-tissue infection, while HA-MRSA is associated with bacteremia and infections in surgical sites. Due to difference

in the genotypic characteristics, CA-MRSA and HA-MRSA can be differentiated rather easily. CA-MRSA tend to be less resistant against antibiotics (i.e. clindamycin susceptible), compared to HA-MRSA isolates, but are known to be more virulent, caused by the possession of genes encoding for Panton-Valentine leucocidin (PVL), a pore-forming toxin. Additionally, the *SCCmec* type is different between CA- and HA-MRSA, where the former typically contain *SCCmec* type IV or V, while the latter possess *SCCmec* type I-III. The combination of the PVL encoding genes lukS-PV and lukF-PV, with *SCCmec* V or IV are reported to have increased pathogenicity, and infection rate, making certain CA-MRSA strains more dangerous than HA-MRSA strains [10]. *SCCmec* typing can be used for epidemiology surveillance of MRSA, classifying strains by the *SCCmec* type (I to IX), along with other genotypic methods introduced in later sections, such as pulsed field gel electrophoresis [15].With the increasing prevalence of MRSA, isolates that are characterised as community acquired, are becoming more prevalent in hospitals, as transmission of these strain types spread in healthcare facilities [24].

2.3.2 Vancomycin resistant enterococci (VRE)

E. faecium and *E. faecalis* are two *Enterococcus* species that are clinically important due to their ability to acquire resistance against vancomycin, as well as being the two common species that cause infections in humans [25]. Enterococci are tolerant to β-lactam antibiotics, and are also intrinsically resistant to many antibiotics, such as semi-synthetic penicillin, clindamycin and cephalosporins. Some strains of VRE are also known to have resistance against last-line antibiotics like linezolid and daptomycin, which limits treatment options for VRE infections [26, 27]. *E. faecium* is known to acquire plasmids, prophages, and genomic islands, among other genetic mobile elements through horizontal gene transfer. This genomic plasticity makes *E. faecium* fit for adaptation to its environment, and is also the reason why they have been successful in evolving to acquire drug resistance, and persistently colonize healthcare facility environment and transmit to susceptible hosts [27]. CDC estimated 54,500 cases of HAI, and 5,400 deaths caused by VRE in 2017 in the USA. Approximately 30% of enterococcal HAI were by VRE (~20,000 infections) in the USA, with the majority identified as *E. faecium* [7].

2.3.2.1 Resistance mechanism against vancomycin in *Enterococcus* species

Vancomycin is a glycopeptide, that is used as the last line of defense against Gram-positive bacteria. It interferes the biosynthesis of peptidoglycan, by forming a complex with the peptide

termini, D-Ala-D-Ala of peptidoglycan precursors. By doing so, peptidoglycan precursors cannot be used as substrates for cell wall biosynthesis (transglycosylation and transpeptidation). Vancomycin resistance in enterococci occurs as the result of an alteration in the peptidoglycan precursor peptide termini, with the replacement of D-Ala with D-Lac or D-Ser, which decreases the affinity by 1000 and 7 folds respectively for vancomycin. This amino acid modification enables cell wall synthesis to continue for bacterial growth. The gene operon that dictates the bacteria's capacity to be resistant can be found in the core genome or on mobile genetic elements. In the latter case, horizontal gene transfers can occur, and pass the resistance gene to enterococcal colonies that were otherwise susceptible to vancomycin. There are nine known gene clusters that cause glycopeptide resistance in enterococci, but the most common types are *vanA* and *vanB* in clinical VRE isolates. The Van gene clusters differ from one another by the physical location (on a mobile genetic element or in the core genome), specific glycopeptides that they are resistant to (vancomycin only or vancomycin and teicoplanin), the degree of resistance, if resistance is induced or constantly expressed, and the type of peptidoglycan precursor produced (D-Ala-D-Lac or D-Ala-D-Ser) [26]. Although less common, *vanA* gene have also been found in other enterococcal species such as *E. avium*, *E. durans*, and *E. raffinosus*. *vanA* genes are associated with inducing high resistance level to both vancomycin and teicoplanin, while *vanB* gene induces moderate to high-level resistance to only vancomycin. The peptidoglycan precursor produced by both *vanA* and *vanB* is D-Ala-D-Lac. *vanA* gene are encoded on *Tn1546*, while *vanB* gene can be found on *Tn1547* or *Tn1549* mobile elements on plasmids or in the chromosome [26, 28].

2.4 Current techniques for bacterial identification and antimicrobial susceptibility testing

As part of the screening process for nosocomial pathogens, a combination of genotypic and phenotypic techniques are used for identification and determination of the level of antimicrobial resistance. Matrix assisted laser desorption ionization – time of flight mass spectrometry (MALDI-TOF MS) recently became the overwhelmingly popular choice for microbial identification due to its ease in sample preparation and rapid turnaround time[29]. Genotypic methods such as polymerase chain reaction (PCR) can identify targeted species and resistance genes, while antimicrobial susceptibility tests (AST) like microdilution inhibition tests can quantitatively determine the level of resistance. Without the knowledge of the antibiotic resistance profile of the isolated microorganism, misuse of antibiotics may occur, where antibiotics may be ineffective in treatment or lead to antibiotic resistance due to unnecessary exposures. Phenotypic tests such as

broth microdilution, disk diffusion test, and agar dilution are the gold standard for antimicrobial susceptibility testing , due to its flexibility, as well as the ease to conduct the tests, in order to determine the therapeutic dose required in infection treatments [30].

2.4.1 Biochemical reaction-based identification systems

The VITEK system (bioMérieux SA, Marcy l'Étoile, France) was first developed in the 1970s, as an identification and AST system, based on a series of biochemical reactions which measures metabolic activities. Some of the biochemical changes that are observed for analysis include carbon source utilization, acidification, alkalinization, enzymatic hydrolysis, effect of inhibitory substances [31]. Other manufacturers have developed similar automated biochemical based identification systems, such as BD Phoenix (BD Diagnostics). Briefly, a 0.5-0.63 McFarland turbidity standard is prepared with 0.45-0.5% saline solution. This is then automatically filled into a card with many wells, and inserted into the system where the card is incubated at 35.5 °C and data is acquired every 15 minutes using an optical system which measures turbidity, color, and colorimetric signals. Results for Gram-positive bacteria can be identified in eight hours or less [31, 32].

AST is a feature of the VITEK 2 system, where antibiotics are tested for MIC. For example, the antibiotics tested for Gram-positive cocci in clusters that suggest *Staphylococcus spp*., are benzylpenicillin, cefoxitin, vancomycin, rifampicin and linezolid. For Gram-positive cocci in chains, suggesting *Enterococcus* or *Streptococcus spp*., MIC against benzylpenicillin, ampicillin, imipenem, vancomycin, linezolid and high-level gentamicin are determined [33]. Based on the optical density of the microwells after incubation with a range of antibiotic doses, MIC is determined for each tested antibiotic. Rapid (4 to 15 hours) identification by VITEK 2 AST have been reported, with accuracies between 94-100%, 90-100%, and 91-100% for *S. aureus*, CoNS, and *Enterococcus* species respectively [32, 34], making it a reliable AST method.

2.4.2 Matrix assisted laser desorption ionization – time of flight mass spectrometry (MALDI-TOF MS)

In recent years, MALDI-TOF MS has become a popular microbial identification technique that is widely used around the world in application for clinical diagnostics. Ease of preparation and use of system, as well as the rapid turnaround time (as fast as under an hour) for results are key reasons for the popularity. MALDI-TOF MS, like other mass spectrometry (MS) techniques,

ionizes analytes into charged molecules, and produces mass spectra, where the intensity is plotted over a range of mass to charge (*m/z*) ratio. MALDI-TOF MS was developed in the 1980s, with the intended use for biological samples, which are larger in size compared to chemical samples traditionally analyzed by other MS techniques. [35]. Identification of the microorganism by MALDI-TOF MS is based on the use of a spectral database, and identifying species with high degree of similarity in the *m/z* profiles observed in the range of 2-20 kDa, which correspond to the ribosomal proteins detection region [35].

The advantages of MALDI-TOF MS is that it is a rapid technique, with simple preparation steps, and can produce accurate results. Despite the high initial cost for the equipment, cost per test is less expensive compared to traditional molecular and immunological based methods, as a result of smaller volume of reagent required. MALDI-TOF MS reported a significantly reduced turnaround time for identification, resulting in faster implementation of appropriate antibiotic treatment, in reference to biochemical assays. In combination with the reduced use of consumables and reagents, the use of MALDI-TOF MS resulted in reduction of overall waste and expense associated with bacteria identification [36]. Unlike techniques such as polymerase chain reaction (PCR) which require highly trained personnel to perform the tests, MALDI-TOF MS does not require extensive training. MALDI-TOF MS is recognized as a culture media independent method; studies have shown that the use of different media such as blood agar, tryptic soy agar, and chromogenic agar were able to obtain the same identification results, because the ribosomal protein expression does not alter from the change in media [37, 38]. In order to obtain high confidence results, it is critical that the sample being analyzed have been well isolated.

In terms of performance, MALDI-TOF MS was reported to have lower error rates at both the genus and species level compared to VITEK 2 [39]. Bruker Biotyper and VITEK MS MALDI-TOF MS systems, manufactured by Bruker Corporation and bioMérieux SA respectively, were able to identify microorganisms at 85 % or better, even when identifying unusual or difficult microorganisms, showcasing comparable, high results between the two manufactures [29, 36, 40]. Despite many advantages, MALDI-TOF MS has a few disadvantages as a bacteria identification system. Due to the reliance on difference in a limited region of the mass spectrum, groups of bacteria that are closely related that have highly similar ribosomal protein mass fragment patterns may be unsuccessfully differentiated [41]. Examples include differentiation between *Shigella spp*. vs *E. coli*, *Streptococcus pneumoniae* vs *S. mitis*, species within the *Enterobacter cloacae complex*,

B*urkholderia cepacian complex*, and *Acinetobacter spp*. [42]. Expanding the *m/z* region of interest to include those outside of the current 2-20kDa region may result in improved identification by MALDI-TOF MS [43]. Unfortunately, not all hospitals can afford a MALDI-TOF MS systems, especially in small hospitals, where funding is limited, and sample throughput is low, and therefore cheaper alternatives that are able to produce similarly reliable results are necessary.

In extension to using MALDI-TOF MS for rapid identification from isolated colonies, interests in direct identification from positive blood culture and urine samples have been increasing, owing to the very little biomass required for analysis, and the urgency of bloodstream and urinary tract infections to be treated [44-48]. In the study by Mitchell and Alby, MALDI-TOF MS and VITEK2 AST used to identify isolates from small colonies obtained on solid agar at minimal incubation for a faster turnaround time. Identification and AST results were obtained within 5 to 12 hours from positive blood culture, with 84-88 % concordant identification and 99 % categorical agreement to standardized protocols for respective techniques from colonies grown at 24 hours on solid media. [45].

Applications in sub-typing, strain typing, and detection of antibiotic resistance has also been reported, although the *m/z* region used are not the same as the MALDI-TOF MS for bacterial identification. Attempts in differentiating antibiotic susceptible from resistant strain have been reported by detecting peaks related to antibiotics and their metabolized products [35, 48, 49]. One report showed that MALDI-TOF MS can detect characteristic peaks that are specific to MRSA and MSSA, and had 84.2 % and 90.9 % accuracy rate respectively [50]. Another reported the capabilities of MALDI-TOF MS to not only identify the organisms, but also detect *vanA* or *vanB* type resistance in *E. faecium* and type the isolates for relatedness from the same spectra [51]. Despite these reports, the use of MALDI-TOF MS for sub-species level discrimination is still in debate. There are contradicting reports of successful and unsuccessful discrimination for strain typing and detection of antibiotic resistant strain from their susceptible counterparts such as MRSA vs MSSA and VRE vs VSE [41, 52, 53]. The discrepant reports maybe be due to differences in, or insufficient representation in the strain types of these organisms. For example, the biomarker (hiracin) identified by Griffin et al (2012) for *vanB*–type VRE was reported to only be found in a subpopulation of VRE. Use of such biomarkers was criticized of having low positive predictive values [54]. Bruker Corporation's MALDI Biotyper released a method for identifying antibiotic resistance in KPC-producing *Klebsiella pneumoniae*, MRSA, and carbapenemase-producing

Bacillus fragilis using MALDI-TOF MS. Although identification of KPC-producing *K. pneumoniae* and caprbapenemase-producing *B. fragilis* were successful, MRSA was identified with low specificity. MRSA detection relied on the detection of a peak related to phenol soluble modulin - mec (PSM-mec) peptide. A large geographical variation in the proportion of MRSA that possess PSM-mec was discovered, explaining the low sensitivity in the MALDI-TOF MS method [52, 55-57]. Mixed results on successful strain typing of organisms for outbreak detection by MALDI-TOF MS proves the need for more extensive studies to fully evaluate its potential [58-60].

2.4.3 16S rRNA sequencing

16S rRNA is a part of the 30S ribosomal subunit, that is ~1500 base pairs. It is the most commonly used housekeeping genetic marker for studying phylogeny and taxonomy, and is considered a reliable molecular marker for identification. 16S rRNA is present in all prokaryotic cells and has conserved and variable sequence regions that evolve at different rates. It is especially useful for identification of organisms that are fastidious and slow-growing such as mycobacteria as it does not rely on the use of culturing. The main limitation is that there are limited resolution, which results in low correct identification, especially at the species level. Bacteria of different species or genera can have high similarity in the 16S rRNA genes, despite having distinct biochemical and overall DNA homology [61]. Srinivasan et al (2015) reported 96 % and 87.5 % correct identification by 16S rRNA sequencing at the genus and species level upon testing 617 samples from a wide range of Gram-negative and -positive organisms [62], while others have reported as low as 62.5 % correct species level identification [61]. 16S rRNA sequencing is therefore not frequently used for routine identification, and is rather used for determining phylogenetic studies, reclassification of taxonomy and identification of new species [61].

2.4.4 Polymerase chain reaction (PCR)

Polymerase chain reaction (PCR) is a genotypic technique that can be used for detection of specific organisms [63], antibiotic resistance [64, 65], or toxins [66] based on the presence of their associated gene. The target gene is amplified using thermocycle, and detected for its presence. PCR is useful in screening for specific pathogens, such as VRE, and MRSA, as it requires the use of specific primers, to detect target gene relating to their antibiotic resistance. For VRE screening, the detection of *vanA* and *vanB* genes are used [67, 68] while *mecA* gene is used for MRSA

detection or methicillin resistance in other staphylococcus isolates [69]. Due to the presence of *mecA* in methicillin resistant coagulase negative staphylococci, PCR primers that detect *orfX* or *nuc* gene, specific to *S. aureus* has also been developed, allowing a one-step detection for MRSA, by simultaneously detecting both genes [23, 70]. Multiplex PCR panels have been developed for detection of multiple microorganisms from respiratory and central nervous system samples [71]. Although rapid resistance gene detection (or the lack thereof), can lead to adjustment in appropriate antimicrobial therapy reflected as the escalation or de-escalation of treatment, the lack of the resistance gene does not necessarily ensure susceptibility to the associated antibiotic. The lack of understanding in correlation between genotypic to phenotypic characteristics in terms of antibiotic susceptibility, therefore requires that genotypic AST should be a supplement to phenotypic test [30]. The time to detection is significantly shorter than culturing methods, and some such as BioFire Diagnostic's panels for pneumonia, meningitis/encephalitis, and respiratory have the capabilities to identify viral infections, which is important in determining whether antibiotics are required for treatment [30]. However, multiplex PCR panels are known to be expensive, making it less accessible for routine diagnostic use [30].

2.4.5 Chromogenic agar

Chromogenic agar is a selective and differential media utilized to screen and identify specific pathogens, including those that are antibiotic resistant. It contains chromogenic substrates, which can only react with specific to enzymes that are possessed by target pathogen(s). Color is produced through enzymatic modification of the chromogen, allowing for visual identification on the surface of the agar plate upon cultivation. Chromogenic agar is selective and differential in nature, as it contains compounds that inhibit the growth of non-target microorganisms that can metabolize the same chromogenic substrate as the target pathogen, and often, selective reagents such as inhibitors and antibiotics are added to prevent growth of non-targeted microorganisms. This reduces growth of background flora, which could otherwise overwhelm target organisms or cause them to be overlooked on generic media. The goal of using chromogenic agar is to make identification of target pathogens easier, which may lead to faster turnaround time for confirmation of pathogen colonization or infection. [72].

Chromogenic agar targeted to detect MRSA was first reported in 2000. MRSA detecting chromogenic agar contains cefoxitin which inhibits the growth of MSSA, thereby leaving positive

growth of MRSA on the agar. Improvement in sensitivity is observed through incubation for 48hours, compared to 24 hours. However, it also known that there is a decrease in specificity, as non-MRSA isolates show positive growth at 48 hours. Improvement in sensitivity for MRSA detection on chromogenic agar can be done using an overnight broth enrichment step prior to inoculation onto the agar, rather than directly from swabs [72, 73]. Chromogenic agar used to screen for VRE contains vancomycin to inhibit growth of VSE. In chromID VRE (bioMérieux SA, Marcy l'Étoile, France), α-glucopyranoside and β-galactopyranoside are utilized to visually differentiate between *E. faecalis* and *E. faecium*, through generation of different colors (green-blue and purple respectively). Positive growth on chromogenic agar aid in confirmation of VRE, especially for *vanB* VRE, since PCR has high false positive rates caused by the possession of *vanB* gene in non-enterococcal species [68].

2.5 Strain typing techniques for surveillance of nosocomial pathogens

As highlighted in previous sections, surveillance and investigations of nosocomial outbreaks are critical in preventing spread of the causative agent, especially of those that are antibiotic resistant strains. Hospitals have different surveillance protocols depending on their size and capacity. Surveillance is important, as it provides a baseline rate, allowing detection of atypical surge in colonization and infection rate, which are referred as outbreaks. Long term trends can be observed, and through surveillance any new or unusual strains identified can be requested for further testing and characterisation. Screenings are commonly used to track colonization and transmission of pathogens within healthcare systems. There are two types of screening: active and passive. Active screening is the act of routinely testing patients for certain organisms such as VRE and MRSA, typically at the time they are admitted, and also on a weekly or bi-weekly basis depending on the protocol set at the facility. Passive screening is the identification of target organisms from sources such as infections, rather than from routine screening. Passive screening is not ideal because it does not allow the hospital to monitor asymptomatic colonization, which can be transmitted to new patients. Continued routine screening can also aid in removing patients from contact precautions once decolonization is confirmed [74]. Since asymptomatic carriers are major reservoirs, acting as the source for transmission of drug resistant pathogens, active screening not only helps non-colonized at-risk patients from increased risk of HAI, but it also reduces healthcare costs in the long run, if outbreaks and transmission can be prevented [75].

Nasal and rectal screenings are common for identification of MRSA and VRE respectively. Environmental screenings can also detect reservoirs of pathogens on abiotic surfaces such as gowns and medical equipment. Once target pathogens are identified, strain typing is required to provide molecular evidence to establish the presence of nosocomial outbreaks, in combination with patient history and space-time epidemiological data. Pulsed field gel electrophoresis (PFGE) is considered the gold standard in strain typing for epidemiological studies due to its high discriminatory capabilities, and its ability to analyze a variety of bacteria by choosing the appropriate restriction enzyme for DNA fragmentation. Other methods have also been developed for strain typing, such as ribotyping, serotyping, *spa* typing and multilocus sequence typing (MLST) but may be more specific to certain microorganisms, and/or less discriminatory compared to PFGE [76, 77]. Whole genome sequencing (WGS) is gaining popularity in the clinical microbiology world, and is thought to be very useful for epidemiological purposes. Due to its high information content and sensitivity, WGS may replace PFGE as the new gold standard in molecular typing in the near future.

2.5.1 Pulsed field gel electrophoresis (PFGE)

PFGE is currently the gold standard in strain typing of a variety of microorganisms including MRSA and VRE. The determination of clonality between isolates by PFGE rely on differences in the band patterns created by fragmented DNA fractions. Briefly, DNA are extracted from cells, fragmented using restriction enzyme, which are then separated within an agarose gel using electrophoresis, based on their molecular weight. The final PFGE band patterns are compared pair-wise and in batches to evaluate the degree of relatedness between isolates. Although it has been considered the gold standard for so many years, there are criticisms to the technique. PFGE is time consuming (upwards of 5 days), labor intensive, utilizes costly reagents, and requires technicians to be highly trained to run the analyses. Perhaps the most concerning point is that data analysis is subjective, and will vary depending on the technician conducting the analyses. The methods' variability between users is a problem, which hinders its use for easy data sharing and comparison between laboratories in large-scale epidemiological surveillances, as it results in a lack of reproducibility. PFGE has demonstrated successful use in surveillance at national (USA and Canada) and international levels, through rigorous method standardization, which included specifying the equipment, restriction enzymes, and reagents to be used. Examples include PulseNet for identification of foodborne pathogens related to multi-state outbreaks, as well as a nationwide

PFGE surveillance in Canada [60]. Ten laboratories across Europe also conducted a study in effort to harmonize PFGE protocols for MRSA strain typing [78]. Despite extended efforts to standardize the method for transferability and widespread comparison of isolate, PFGE still hold an inherent limitation. As genetic mutations and transfer of genetic elements occur in isolates and between cells over time and in varying environment, there are cases in which phenotypically different isolates may have the same PFGE band characteristics. In other cases, clonal strains may have different PFGE band patterns, due to a mutation that results in added or reduced fragmentation, or generation of different molecular weight fragments [79, 80]. Microorganisms that have high genome plasticity, such as enterococci make them more susceptible to such changes, and the reliability of strain typing by PFGE is reduced for these organisms. These genetic tendencies that are unique to each microorganisms should be understood and considered while determining strain types and isolate relatedness using PFGE. For example, sample collection dates should be taken into account during analysis, especially for such organisms with high genomic plasticity, which may observe higher degree of genetic change compared to other microorganisms that are more genetically conserved in the same time frame [60, 81].

Other genotypic methods, such as multi-locus variable number of tandem repeats analysis (MLVA), spa typing (for MRSA), and MLST have been compared against PFGE, challenging to replace it as the gold standard over the years. However, reports varied in degrees of success dependent on the target pathogen. The fact that PFGE uses the genome in it entirety as fragmented patterns, whereas these other methods utilize small portions of the genome, enabled PFGE to remain as a superior technique over these other genotypic methods [60].

2.5.2 Whole genome sequencing (WGS)

Whole genome sequencing in its first phase was a time-consuming task. However, with the development of next generation sequencing (NGS) instruments, high-throughput genome sequencing became possible, allowing the entire genome of bacteria to be sequenced within a day. In combination with improved computing and processing power, time required for sequencing and data analysis has been significantly shortened. WGS is able to observe differences as small as one nucleotide difference between isolates, showcasing extremely high resolution, providing an enormous amount of data. WGS has been gaining popularity for epidemiological typing, and many have compared WGS against current techniques such as PFGE and MLST, in providing genotypic

data for outbreak investigations, reporting high discriminatory power and success in identifying clonal strains [81-83]. The main advantages of WGS over traditional genotypic typing techniques, are that the resolution is higher, and is able to provide more information from the genome, such as possession of antibiotic resistance genes, virulence genes, plasmids, as well as its ability to track phylogenetic evolutions of the bacteria. The ability to look at the entire core and accessory genomes, rather than specific regions of it also expands the genotypic content that can be analyzed to answer strain typing and other microbiological questions [84].

There are different sequencing technologies that provide long or short reads. The second generation sequencing platforms, such as Illumina MiSeq are known to be highly accurate with low error rate, but has the disadvantage of producing short reads, which make it difficult to assemble complex genomic regions [84]. Illumina sequencers sequence by synthesis, determining the genetic sequence by using fluorescently labeled nucleotides that bind with DNA fragments, known as clusters. Fluorescent labeled nucleotides are added to flow cells where they are incorporated into the DNA fragments, and captured as images, before fluorescent colors are cleaved off, and the next fluorescent labeled nucleotides are incorporated and repeated. In contrast, third generation sequencing platforms like Pacific Biosciences' Sequel system is based on single-molecule real time (SMRT) sequencing. This method detects the colors of the labels that cleave off of fluorescently labeled nucleotides, as they get incorporated into the DNA strand. Light pulses are recorded in real time, to determine the nucleotide sequence, as a continuous long read (> 20kb). The disadvantage of the long read sequencing is that the error rate is higher than short read sequencing methods, as a result of errors accumulating from detection of nucleotides that are not being incorporated but dwell in the active site of DNA polymerases long enough for detection [84]. Long read sequencing is suitable for *de novo* assembly. Genomes are assembled using one of various analysis tools available as web-based or command line. Hybrid assembly enables genome data acquired from both long and short reads together, to further improve the assembly, in comparison to when data from each is assembled independently. Short read sequences correct errors made in long read sequences, and overall, the repeat resolution increases and gaps are filled in the assembly [84].

For strain typing, different genome analysis methods can be employed, such as core genome single nucleotide polymorphism (cgSNP), core genome MLST typing (cgMLST), whole genome MLST, and clustered regularly interspaced short palindromic repeats (CRISPR) [81, 82, 85]. Core

genome MLST typing (cgMLST) and SNP analyses were compared for strain typing of VRE (*E. faecium, E. faecalis*), resulting in comparable results between the two analysis methods [81, 82]. WGS data can be used to determine the MLST sequence type *in silico*, using housekeeping genes and identifying the alleles on each gene. Conventional MLST only utilizes the data associated with the housekeeping genes, while cgMLST and wgMLST uses the common genes identified in the core and pan genome respectively between isolates, utilizing more of the genomic data to compare and determine differences between isolates in question for strain relatedness [84].

WGS data provides insight on the presence of genes related to antimicrobial resistance in isolates [14, 79, 86]. Resistance genes can be identified using tools such as ResFinder [87], among other web-based tools. Studies comparing WGS-based antimicrobial-resistance detection against phenotypic AST in 155 *E. coli*, 332 *Salmonella* Typhi and 271 *Salmonella* Paratyphi samples, reported that detection of known genes related to antimicrobial resistance was a robust approach for monitoring antimicrobial resistance in bacteria, for surveillance purposes. WGS can provide presumptive detection of antibiotic resistance, via detection of known drug-resistance genes and mutations, before phenotypic results are available [88-90]. Additionally, information regarding the presence of virulence genes, plasmids, and phages can be determined from web-based tools such as VirulenceFinder [91], PlasmidFinder [92], and PHASTER [93, 94]. Detection of plasmids are important, as antimicrobial resistance genes and virulence genes could be present on plasmids.

It is estimated that with the reduction in cost, and increased speed in analysis via automated, streamlined analysis, WGS for epidemiological typing can provide answers as fast as 48hrs from obtaining the sample, making it more available for routine use. WGS has been used to track the transmission of pathogens such as MRSA, VRE, *Listeria monocytogenes*, group A Streptococcus and *Acinetobacter baumanii* in hospitals and their surrounding communities [95-100], proving to be a very valuable technology for infection prevention and control, and surveillance at both local and global scales [101]. In order for WGS to be routinely used in clinical laboratories, standardization is required, as there are various sample preparation, data acquisition, genome assembly and analytical methods being used currently by many researchers [84].

2.6 Whole-organism fingerprinting techniques: potential in rapid identification and strain typing applications

Whole-organisms fingerprinting techniques acquire information from whole-cells in a non-destructive manner. These include spectroscopic techniques, such as infrared spectroscopy, Raman spectroscopy and nuclear magnetic resonance (NMR) spectroscopy. Spectra acquired by these techniques are made up of biochemical signals obtained from all components of the bacteria (e.g. proteins, carbohydrates, lipids, nucleic acids), rather than specific components such as ribosomal proteins in MALDI-TOF MS, or the DNA for genotypic methods. Non-destructive analytical techniques therefore allow microbial samples to be analyzed in its native form. Whole-organism fingerprinting techniques are typically armed with multivariate statistical tools such as hierarchical cluster analysis (HCA) and principal component analysis (PCA), for applications in discrimination or classification . Like WGS, where the genomic data can be applied to various analysis for mining a range of information regarding the microorganisms, the application of various multivarious statistical analysis methods allows the use of one spectrum for species identification, but also for characterization and determine spectral relatedness between microbial samples at the subspecies level (e.g. serotyping, antibiotic resistance, strain typing).

2.6.1 Vibrational spectroscopic techniques

Vibrational spectroscopic techniques include infrared (IR) and Raman spectroscopy. Both techniques measure changes in vibrational energy, when samples interact with an energy source. IR and Raman spectroscopy are complimentary techniques; infrared observes stretching and bending vibrations of functional groups, whereas Raman observes polarizability. IR spectroscopy in the mid-IR region (4000-400 cm^{-1}) is by far the most commonly used to study bacteria, although there are exploratory studies reported for near-IR [102-108] and Raman spectroscopy [109-118] as well.

2.6.1.1 Mid-infrared spectroscopy in clinical diagnostics and bacteria identification

Since the 1950s when IR spectroscopy was first used as a technique for bacteria identification, many studies have demonstrated IR spectroscopy's ability to discriminate bacteria at various taxonomic levels including sub-species level, such as serotypes, toxin types, and antibiotic susceptibility for both Gram-positive and Gram-negative bacteria. The development of Fourier-transform infrared (FTIR) spectroscopy allowed the technique to have excellent sensitivity and

reproducibility, while shortening the spectral acquisition time. Combined with the increased processing power of computers, FTIR spectrometers gained the ability to obtain high quality, highly reproducible bacterial spectra rapidly (~1 minute), and non-destructively, with minimal, if any sample preparation at all [119].

Like the MALDI-TOF MS, identification by FTIR spectroscopy requires a spectral database developed from spectra collected from pure and well characterized microbial isolates. The ability to identify the bacteria accurately is therefore dependent on the size of the spectral database(s) and its ability to represent the diversity observed at the genus and species level by the current taxonomic classification system. The FTIR spectrum of a microbial isolate reflects the metabolic and structural characteristics at the time of spectral acquisition. This entails that the spectral characteristics of an isolate are affected by culturing conditions such as growth media type, temperature and incubation time [120-122]. To identify unknown microorganisms from an FTIR spectral database, one of two things must occur: 1. isolates much be cultured only using the culture medium that is present in the spectral database, or 2. develop spectral databases using FTIR spectra acquired from microorganisms grown on many different culture media, so that the spectral database is effectively media independent.

There have been many studies showcasing the potential use of FTIR spectroscopy as a diagnostic technique, using various spectral acquisition methods, such as transmission, and attenuated total reflectance, using single element detectors or focal plane array detectors (Table 2.1). The results from different reports all demonstrate that IR spectra contain sufficient chemical information and sensitivity for bacterial discrimination even at sub-species level. In the following subsections, different FTIR spectral acquisition modes are briefly reviewed.

Table 2.1 List of publications on microbial discrimination using FTIR spectroscopy (2011-2019)

Tested Organism	**FTIR Spectral Acquisition Mode**	**Reference Method**	**Reference**
***Candida* species**	ATR	ChromAgar Candida, PCR (ITS1 and ITS4)	[123]
Foodborne Pathogens (*Salmonella, Listeria, Shigella*)	ATR		[124]
***Klebsiella* capsule typing**	ATR (Dry)	MLST, PFGE	[125]
***Acinetobacter baumannii* strain typing**	ATR (Dry)	PFGE, MLST, carbapenem-hydrolysing class D β-lactamase-CHDL	[126]
Salmonella serogroups and serotypes	ATR (dry)	serotyping	[127]
Shigella* spp. vs *E. coli	ATR (Wet)		[128]
Yeast identification	ATR (Wet)	MALDI-TOF MS, gene sequencing on D1/D2 or ITS region	[129]
***Proteus mirabilis* strain typing**	ATR (Wet)	O-specific polysaccharide of lipopolysaccharides	[130]
***Campylobacter* species**	FPA-FTIR imaging transmission (ZnSe)		[122]
***E. coli* O157:H7 strain typing and toxin-based typing**	reflectance microscope	MLVA and PFGE	[131]
Candida species Differentiation (*C. albicans, C. glabrata, C. krusei*) ChromAgar	Specular reflectance	sequencing (D1/D2 region of the ribosomal 26S gene, NL1 and NL4 primers)	[132]
CA-MRSA vs HA-MRSA	Transmission	PFGE	[133]

Abbreviations: **PCR** - polymerase chain reaction; **MLST** – multilocus sequence typing, **PFGE** – pulsed field gel electrophoresis, **CHDL** – carbapenem-hydrolyzing class D beta-lactamases, **MALDI-TOF MS** – matrix assisted laser/desorption ionization time of flight mass spectrometry, **ITS** – internal transcribed spacer, MLVA – multiple locus variable-number tandem repeat analysis , CA-MRSA – community acquired methicillin resistant *S. aureus*, HA-MRSA – hospital acquired methicillin resistant *S. aureus*

(continued)

Tested Organism	FTIR Spectral Acquisition Mode	Reference Method	Reference
***Salmonella* serorgoups and serotypes**	Transmission	serologic identification (agglutination with specific serums)	[134]
Filamentous fungi identification	Transmission (Silicon)	DNA Sequencing	[135]
***Candida* species identification**	Transmission (Silicon)	WGS	[136]
Klebsiella strain typing	Transmission (silicon)	WGS (SNP analysis), MLST	[137]
GN bacilli (*P. aeruginosa, K. pneumoniae, E. cloacae, A. baumannii*) – Strain Typing	Transmission (silicon)	MLST, PFGE	[138]
Yeast identification	Transmission (ZnSe)	PCR, RFLP of 5.8 R-ITS region	[139]
***Trueperella pyogenes* identification**	Transmission (ZnSe)		[140]
Food-related bacteria	Transmission (ZnSe)	16S rRNA sequencing, PCR	[141]
***Listeria* species identification**	Transmission (ZnSe)	PCR (16S-23S IGS region)	[142]
***Bacillus* species and subspecies**	Transmission (ZnSe)	pacC nucleotide sequences, presence of toxins,	[143]
UPEC strain typing	Transmission (ZnSe)	MLST	[144]
Carbapenem-resistant *Acinetobacter baumannii* strain typing	Transmission (ZnSe)	PFGE, MLST, sequence groups, blaoxa-51-like	[145]
***S. aureus* sub-typing**	Transmission (ZnSe)	PFGE, spa typing	[146]
***Klebsiella oxytoca* strain typing**	transmission (ZnSe)	MLST, PFGE, WGS (SNP)	[147]
***Vibrio parahaemolyticus* sub-typing**	Transmission (ZnSe)	PCR	[148]
ESBL+/- *E.coli*	Transmission (ZnSe)	Disk diffusion test	[149]

Abbreviations: **WGS** – whole genome sequencing, **SNP** – single nucleotide polymorphism, **MLST** – multilocus sequence typing, **PFGE** – pulsed field gel electrophoresis, **PCR** - polymerase chain reaction; **RFLP** – restriction fragment length polymorphism, **ITS** – internal transcribed spacer, **MALDI-TOF MS** – matrix assisted laser/desorption ionization time of flight mass spectrometry, **IGS** – intergenic spacer, **UPEC** – uropathogenic *Escherichia coli*, **ESBL** – extended spectrum beta-lactamse

2.6.1.1.1 Transmission FTIR spectroscopy

Pioneering studies on bacteria identification by FTIR spectroscopy was done in transmission mode, and hence transmission has been the most frequently used spectral acquisition mode for bacteria discrimination and identification, as seen in Table 2.1. Samples are prepared for spectral acquisition by deposition of colonies (directly from agar plates, or after cell suspension or cell washing) onto IR-transparent material such as zinc-selenide (ZnSe), calcium fluoride (CaF_2) or silicon. Under controlled conditions in a low-temperature oven, samples are dried, and simultaneously may also be deactivated using disinfectants, if desired, prior to spectral acquisition. The disadvantage of acquiring IR spectra by transmission mode is that samples are limited by thickness and concentration, to ensure that the absorbance level is not off scale [119].

2.6.1.1.2 Attenuated total reflectance FTIR spectroscopy

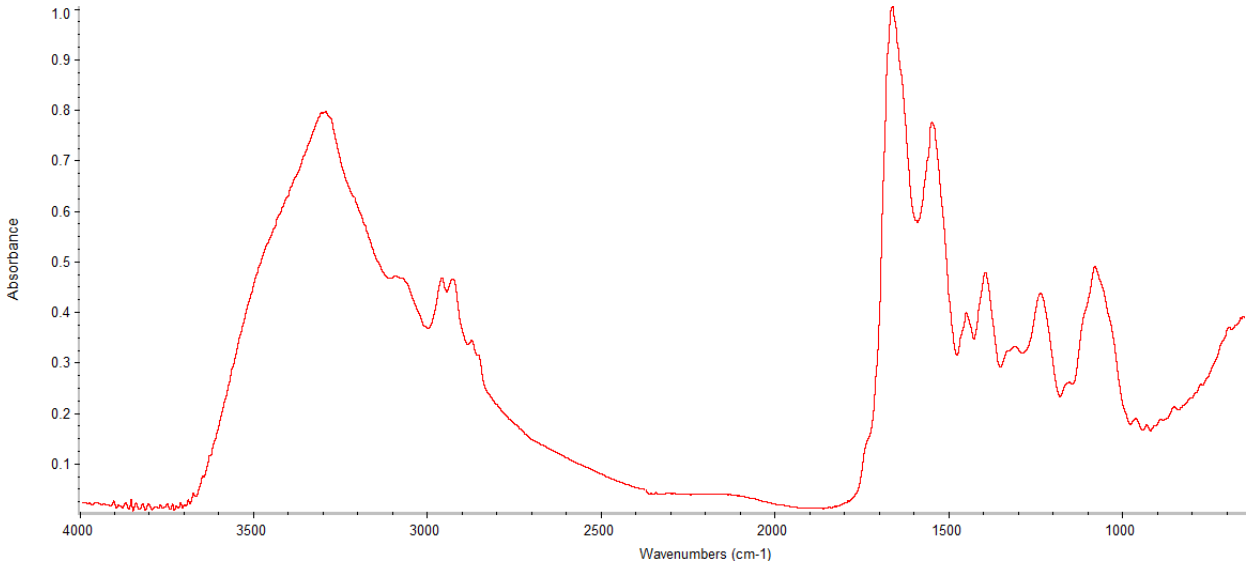
Attenuated total reflectance (ATR)-FTIR spectroscopy utilizes the principle of total internal reflectance of an infrared transmitting crystal (i.e. ZnSe, Ge, diamond, Si) to produce an evanescent wave perpendicular to the propagating IR beam undergoing internal reflection within the crystal. The evanescent wave is emitted into the sample when infrared light internally reflects within the crystal. The evanescent wave interacts with the sample placed on the surface of the crystal, and is attenuated by the sample. The evanescent wave is only able to penetrate between 0.5 to 2 microns into the sample, due to its wavelength dependence and exponential decay in energy. This makes spectral acquisition by the ATR mode immune to sample thickness variability, as long as there is sufficient contact between sample and the ATR crystal. This enables spectral acquisition from solid, semi-solid and liquid samples that would typically be too thick, or too concentrated for analysis by transmission and transflection modes. A few research groups acquired ATR-FTIR spectra by direct application of microorganisms onto the ATR detector from solid media [123, 128, 129, 150, 151], while others made bacteria containing solution, or created dried films on the ATR surface prior to spectral acquisition [126, 152, 153]. The differences in sample preparation does not seem to affect the spectral discriminatory capabilities, however, drying the samples at room temperature directly on the ATR crystal is a time-consuming step, that acts as a bottleneck in a technique that can acquire FTIR spectra in as little time as ~ 1 minute per spectrum.

2.6.1.1.3 Transflection FTIR spectroscopy

Compared to transmission and ATR modes, spectral acquisition by transflection acquisition mode is less studied. Even in the past decade (Table 2.1), only two studies reported the use of transflection for studying microbial discriminatory capabilities. Transflection-FTIR spectroscopy utilizes an IR-reflective substrate, such as low-emissivity (low-e) glass, or slides coated with silver or gold, rather than IR-transparent materials used for transmission spectral acquisition. E-glass is an optically transparent, low-cost material, made through the deposition of thin layers of Ag/ZnO on glass microscope slides, and can be used as single-use disposables [154, 155]. In transflection, the IR source enters the sample, reflects off the reflective surface, and passes through the sample again on its way back to the detector. This technique essentially doubles the pathlength compared to transmission mode, where the IR source only passes through once before reaching the detector. The disadvantage of transflection FTIR is that spectral distortions are more prominent as a result of interferences and reflections caused by change in refractive index at the air-sample interface and sample-reflective surface interface [154-156]. Electric field standing wave effect creates spectral artefacts resulting in spectral distortions. This effect is a result of the creation of sinusoidal standing-wave in the electric field, with a node at the surface of the reflective metal surface and the deposited sample, through the interference between incident and reflected waves. The refractive index of the sample, and the wavelength of the incident light determines the node-antinode spacing [157]. Filik et al. (2012) reported that there is a non-linear relationship between the absorbance and sample thickness in transflection, which was not observed in transmission data, which was determined to be caused by the combined effects of electric field standing wave, and internal reflection from the top of sample surface to air [156]. However, it was reported by Lee (2017) that the effects are not dependent on the sample thickness, but on the interference and reflections that occur at the interface between air and the sample [158]. According to Lee, and other investigators, these spectral distortions as a result of change in refractive index is not a problem that is unique to transflection, and exists in other spectral acquisition modes like transmission, although the observed effect is much stronger in transflection mode [158, 159]. Another concern with transflection-FTIR spectroscopy is that with non-uniform sample thickness, it may suffer from spectral distortions caused by optical phenomena such as resonant Mie scattering [159, 160]. It has also been reported that with consistency in sample thickness, and spectral pre-processing, the effects of electric field standing wave can be minimized [161]. It has

been previously suggested that transflection-FTIR spectroscopy can be used for qualitative but not for quantitative analyses [159]. Up to this point in time, only three studies were reported using reflectance spectral acquisition mode, where one was using specular reflectance (also known as transflection) for microbial studies using FTIR spectroscopy: discrimination between three Candida species [132] , diffuse reflectance for discriminating between *E. coli*, *S. aureus* and *C. albicans* [162] and lastly, reflectance FTIR microscopy in sub-typing *E. coli* [131].

Figure 2.1 Example of a transflection FTIR spectrum of thinly dried bacterial colony



2.6.1.1.4 FTIR imaging and mapping

Focal-plane-array (FPA) - and linear array (LA) – FTIR spectroscopy are used for FTIR imaging and mapping. They are techniques that utilize multiple detectors that are aligned side by side (LA; in a line), or in a grid (FPA i.e. 16x16, 32x32, 64x64, or 128x128 grid detectors). FPA-FTIR spectrometers acquire hyperspectral FTIR images, where each pixel of the FTIR image contains the IR spectrum acquired from its spatial position. The same image can be acquired by a LA-FTIR spectrometer, or even a single element detector spectrometer, but would require much more time, as the number of spectra acquired at once differ substantially. FPA-FTIR spectrometers are often equipped with a microscope, allowing micro-images to be acquired, with spatial resolution as low as ~5 µm across a surface area of 0.3mm x 0.3 mm. FTIR images can be acquired by transmission, transflection or ATR mode, by utilizing the appropriate substrates and accessories.

Extended details on the theoretical aspects, instrumentation, sample preparation and spectral acquisition methods have been explained by [163]. Some examples of studies conducted on bacteria using FPA-FTIR spectroscopy include differentiation between *Campylobacter jejuni* and *C. coli* [122], differentiation between group I and group II *Clostridium botulinum* [164], and the antimicrobial activity of select disinfectants on MRSA [165].

2.6.2 Raman Spectroscopy

Raman spectroscopy is a vibrational spectroscopic technique that compliments infrared spectroscopy, producing spectra based on the observations of Raman inelastic scattering. The scattering relies on changes in polarizability of the functional groups, in particular, non-polar functional groups such as C-C and S-S have strong peaks. Contrary to FTIR spectroscopy which has a very strong absorbance from water, Raman spectroscopy has a low water signal, making it a preferable technique for biological samples, which tend to contain high water content. Due to the weak Raman scattering effect (one in 10^8 photons), longer acquisition time is required to increase the sensitivity, compared to FTIR spectroscopy [110, 113]. Lasers with excitation in the range of infrared to UV have been used in Raman spectroscopy, and within each range (IR, visible and UV) of the electromagnetic spectrum, there are observable Raman scattering peaks associated to biomolecules that can be found in bacteria. For example, UV excitation obtain signals from DNA, RNA and aromatic amino acids, and visible excitation observes signals from chromophores such as cytochromes. However, there are also problems associated with using high-energy excitation lasers. The use of UV excitation lasers has been reported to cause photochemical degradation of the sample, and hence is not recommended for biological samples. Additionally, the lasers that utilize visible and near-infrared excitation can cause fluorescent emissions from samples. Fluorescent emissions have strong and broad peaks, which would overlap and interfere with the samples' Raman scattering peaks [113, 166].

There are two types of Raman spectroscopy techniques that are used in bacteria identification. The first is surface-enhanced Raman spectroscopy (SERS), and the second is Raman microspectroscopy. SERS is known to enhance the Raman scattering process by 10^3-10^6 folds, with the use of a roughened or nanostructured metal, typically gold or silver. The enhanced signal comes from the resonance that occurs between the surface plasmons of the metal surface and the incident and scattered radiation fields. In addition to increasing the signal from samples, SERS

uses lasers that have vibrational excitations, and therefore prevents damage to the analyte. Additionally, SERS has a fluorescence quenching effect, which provides additional information about the bacteria samples being analyzed [113, 115, 166]. The disadvantage of SERS is that the substrate used, as well as its size and shape, may distort the Raman spectra, making it incompatible for combined analysis with spectra obtained by the standard Raman spectroscopy, and even spectra obtained from SERS that utilize different substrates [166]. An example of successful use of SERS for microbial discrimination includes successful species discrimination of 9 staphylococcal species using machine learning algorithms [167].

Raman microspectroscopy is a similar concept to FTIR microspectroscopy. It incorporates a microscope, to observe a small surface area, and increase the spatial resolution where the spectral measurements are obtained. In the case of Raman microspectroscopy, the spatial resolution can be improved to as high as ~1μm, which provides Raman spectroscopy with the potential to obtain spectra from single bacterial cells and microcolonies [110]. This is an advantage in rapid bacteria identification, as it holds the potential to reduce the culturing/incubation time required, prior to spectral acquisition for identification. Maquelin et al (2000) demonstrated that bacteria differentiation was achieved using Raman microspectroscopy, by acquiring spectra from microcolonies (defined as grown for 6 hours from inoculation) directly from solid media [111]. In another study, clinical isolates from UTI patients were differentiated at both species and strain level using SERS alone [118] and in combination with a microscope [113]. A retrospective study on nosocomial *E. coli* outbreak was conducted using Raman spectroscopy [117], showcasing Raman spectroscopy's potential in strain-typing. Some other studies that utilize Raman spectroscopy for bacteria identification include species level analysis on *Mycobacterium* species [116], and a feasibility study for rapid identification of positive blood culture samples using SERS [115]. Spectral acquisition from single cells from dried samples were acquired, and demonstrated the ability to discriminate Raman spectra between 9 genera, using machine learning techniques [168].

2.6.3 NMR spectroscopy

NMR spectroscopy is a technique traditionally used in chemistry, physics and biochemistry, mainly for molecular structure elucidation, amino acid sequence in proteins, reaction rates and dynamic studies. Recently, it has been gaining popularity as an analytical technique to study

biological materials, such as food and beverages, biofluids, tissues and whole cells [169-173]. NMR has the ability to identify the presence of specific biomolecules from its spectrum, achieving higher chemical specificity compared to FTIR and Raman spectroscopy.

Proton (^{1}H) is the most commonly used nucleus in NMR spectroscopy due to its highest natural abundance (99.98 %), and high gyromagnetic ratio (267.51γ). These two factors allow for acquisition of high quality spectra in a relatively short amount of time. Other commonly used nuclei in NMR spectroscopy include ^{13}C, ^{15}N, ^{31}P, and ^{2}H. Phosphorus (^{31}P) is another nucleus that has high abundance (100 %) and relatively high gyromagnetic ratio (108.29 γ), like ^{1}H. Phosphorus NMR enables spectral acquisition and study of phosphorous containing biomolecules, such as lipids, nucleotides, proteins/peptides and carbohydrates without the need for isotopic labelling. It is particularly useful in studying phospholipids, and their structure in lipid membranes. Chemical shift observed in ^{31}P NMR spectral acquisition in solution state is dependent on the solvent used to dissolve the samples. The addition of Cs-EDTA (cesium ethylendiamine tetraacetic acid) eliminates line broadening that would otherwise be present due to the presence of paramagnetic ions [174]. Phospholipids can also be dissolved in aqueous solvents through the addition of cholate [175]. Bilayers in aqueous solvent, and inverse micelles in an organic solvent both cause line broadening in the ^{31}P spectrum. Stable lipid micelles are formed when the solvent contains compounds like cholate which possess both nonpolar and polar groups.

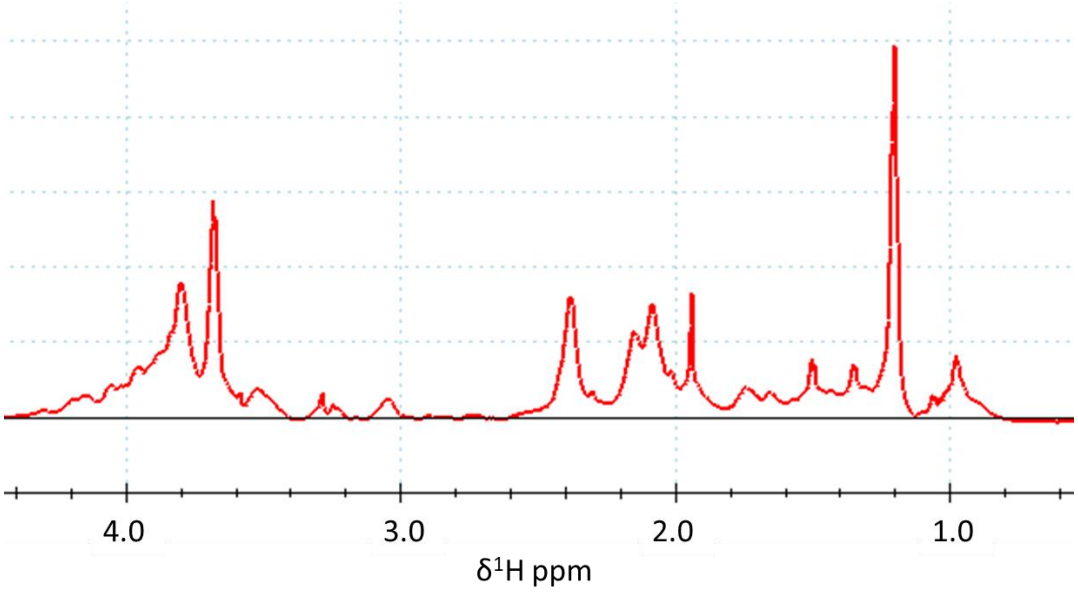
2.6.3.1 High-resolution magic-angle-spinning (HR-MAS) NMR spectroscopy

HR-MAS NMR spectroscopy is a technique that was developed to acquire high resolution NMR spectra that are comparable in spectral quality to those acquired by solid-state and liquid-state state NMR experiments, from semi-solid or heterogenous samples. Many biological samples require the use of HR-MAS NMR spectroscopy, as they are very heterogenous in nature, composed of molecules with varying sizes, shapes, and mobility. In solid-state and liquid-state NMR spectroscopy, such heterogeneity within samples lead to line broadening caused by chemical shift anisotropy and dipolar coupling. In HR-MAS NMR spectroscopy samples are rotated at the magic angle (~54.74°) relative to the external magnetic field to suppress the effects of dipolar coupling and chemical shift anisotropy. This magic angle was determined from the following equation: $\frac{3\cos^2\theta_m - 1}{2} = 0$ where $\theta_m = \sim 54.74°$ [176]. The required sample volume is very small (approximately 100µl), and it is spun at a low speed (up to a few kHz), in a rotor. This allows

NMR spectra to be acquired from intact whole cells and tissues, without damaging them during spectral acquisition, while spinning fast enough that the spinning side bands are not overlapping with spectral peaks derived from samples [177]. Previous studies reported the use of HR-MAS NMR spectroscopy to study bacteria, and showed that the survival rate of bacteria within a sealed NMR insert after two hours was higher than 93 % [178], thereby confirming that the NMR spectra acquired were of live cells. According to Li, HR-MAS NMR spectroscopy is only able to observe biomolecules that are in abundant concentration and are mobile within the bacteria [179].

One dimensional NMR spectroscopy is a commonly used spectral acquisition technique, largely due to its simplicity compared to multidimensional experiments. For bacteria and other biological samples that are comprised of many different compounds, the one dimensional spectrum is comprised of many overlapping peaks in the ^{1}H NMR spectra, and therefore peak assignment and identification of biomolecules is not a simple task. While some metabolites have unique peaks that do not overlap with other metabolites, some require 2D NMR experiments to resolve the peaks from other metabolites for definitive identification [180]. Similarly to FTIR spectra, ^{1}H HR-MAS NMR spectra of biological samples serve as spectral fingerprints that can be acquired rapidly, and provide important chemical information that can be analyzed using multivariate statistical analysis methods for global differences between samples [179]. Multidimensional NMR experiments (2D, 3D and higher dimensions) use and correlate between one or more nucleus. Multidimensional NMR experiments are used to for structural biology, and is useful in identification of metabolites and structural compounds [179]. Compared to 1D NMR experiments, multidimensional NMR experiments are complex, with many parameters to consider such as pulse sequence, rotor geometry and how fast the samples are rotated along the magic angle. Li reported the importance of rotor synchronicity, as it affects the intensity or the detection of signals when not aligned [178].

*Figure 2.2 Example ^{1}H HRMAS NMR spectra of bacteria (*E. faecium*) in spectral region δ = 0.8-4.2ppm*



2.6.3.2 NMR applications in microbiology and metabolomics

There are a limited number of studies that have utilized NMR spectroscopy for microbial discrimination and strain typing. This is most likely due to the high capital expense associated with NMR spectrometers, lack of automation in set-up and lack of NMR laboratories that allow spectral acquisition of live pathogenic microorganisms. Additionally, at the current state, the NMR instrumentation is large, and requires a dedicated space. Miniaturization of the NMR spectrometer and ease of use (through automation) may decrease the hurdles for routine use. Proof of concept papers demonstrated the utility of solution-state NMR spectroscopy in combination with multivariate statistical analysis methods for the discrimination of microorganisms, for example, between *Enterococcus, Streptococcus* and *Staphylococcus* isolates, as well as *Candida species* including two closely related, *C. albicans* and *C. dubliniensis*, with 86-96 % concordance with their respective identification reference methods [181-184]. Additionally, by acquiring spectral data from the broth media after cultivation of microorganisms, differentiation between pathogens could be successfully achieved based on their exo-metabolomes [185, 186]. T2 Diagnostics is a company that developed a benchtop assay that combines NMR spectroscopy and PCR for direct detection of pathogens such as *Candida species* and drug-resistant Gram-positive and Gram-negative pathogens *(i.e. E. faecium, S. aureus, Klebsiella pneumoniae, Acinetobacter baumannii,*

Pseudomonas aeruginosa, Enterobacter species, E coli) from whole blood samples. By utilizing genes specific to target pathogens, the amplified DNA are attached to magnetic beads, which are detected with NMR. It is reported to obtain identification of pathogens, and their antibiotic resistance (via detection of genes) within 3-5 hours from concentration as low as 1CFU/ml [187-189].

Multidimensional HR-MAS NMR experiments also showcased potential in microbial studies. Serotyping was achieved by identifying the polysaccharide moieties on cell wall structures of live *Neisseria meningitidis* and *Candida albicans*, glycans on *Campylobacter species* and O-specific polysaccharides from *Yokenella regensburgei* [190-198]. Gudlavalleti et al reported that the peaks observed in purified capsular polysaccharides could be observed in ^{1}H HR-MAS NMR spectra acquired from the whole cell, suggesting that direct analysis may be able to identify potential biomarkers that are associated with virulence [195]. Maes et al. studied the phenotypic differences between *C. albicans* serotype strains using 2D HR-MAS NMR spectroscopy, reporting that whole cell analysis by HR-MAS NMR spectroscopy was in concordance with in vitro analyses of the polysaccharide structures [198].

NMR spectroscopy is a popular technique for metabolomics due to its non-destructive and non-targeted nature, as well as its ability to acquire data from very small sample volume/size. It has been suggested by many that NMR metabolomics can be applied for drug discovery, studying drug-cell interactions, understanding the effects of gene mutations, mechanisms in pathogenicity, and serotyping/cell surface characteristics [191, 195, 199]. NMR studies on bacteria metabolomics used to be conducted using solid-state spectrometers, upon isolation of molecules of interest (e.g. cell wall, peptidoglycan, etc.) [197, 200]. More recently, there have been metabolomic studies on whole-cell bacteria and yeasts by using HR-MAS NMR spectroscopy, observing the effects of aging, nutrient variation, osmotic differences, drugs and gene mutations [191, 196, 199, 201]. Bundy et al (2005) reported that metabolic extracts from *Bacillus cereus* strains originated from different ecotypes could be differentiated by using NMR spectroscopy in conjunction with chemometrics although there were no significant differences between the samples' genomic DNA and detection of gene related to toxin production and virulence factors present [202]. By acquiring FTIR and HR-MAS NMR spectra from live microbial cells taken directly from agar plates, it allows us to obtained and characterize metabolic information about the strains in its native state.

The two techniques compliment each other, allowing us to further understand the biochemical differences that are reflected in the spectra between microbial isolates.

Whole-organism fingerprinting techniques have strong potential in bacteria identification and discrimination, even at the sub-species level. Non-targeted metabolomics by NMR spectroscopy can help identify key metabolites that are contributing to the observed biochemical differences between strain types by FTIR spectroscopy. Whole organism fingerprinting techniques can take advantage of these microbial characteristics to observe enhanced differences amongst strain types, and possibly differentiate antibiotic resistant from susceptible species through use of selective or differential media. Correlation between genotypic, phenotypic and spectroscopic techniques would aid in better understanding the relationship between genes present in the bacteria with their metabolic expression, allowing for a more comprehensive understanding of the differences at the sub-species level, or between strain types of microbial species.

2.7 Conclusion

Strain type characterization and determination of isolate relatedness for outbreak detection currently relies on the use of PFGE. Furthermore, strain typing is only done once isolates have been identified to be one of the pathogens of interest, through multiple genotypic and phenotypic methods such as MALDI-TOF MS, PCR and automated biochemical tests like VITEK 2 (AST), which also require a minimum of two days. FTIR spectroscopy in conjunction with multivariate statistical analysis methods can provide species identification and strain relatedness at a fraction of the cost, and time compared to current techniques, providing a means for prospective strain typing method, thus providing a rapid and simple, alternative approach to PFGE. This could benefit infection prevention and control in minimizing the transmission and outbreaks caused by antimicrobial-resistant pathogens such as MRSA and VRE. Although it is expected that WGS will become more available for routine use in the near future, there is a lack of understanding in the relationship between genotypic and phenotypic characteristics of microorganisms, which means that phenotypic tests cannot be omitted completely out of the microbial diagnostic routine. In a time where WGS is gaining momentum and popularity for routine applications in clinical diagnostics and outbreak surveillance, whole-organism fingerprinting techniques can provide valuable information on the spectroscopic characteristics based on the metabolic and structural composition of microbial samples. Tranflection FTIR spectroscopy was identified to be less commonly used in studying microbial discriminatory applications by FTIR spectroscopy, and thus would benefit a series of experiments that address whether there are inherent limitations in the transflection mode that prevent its use for microbial diagnostic. Additionally, the combined use of transflection FTIR spectroscopy and HR-MAS NMR spectroscopy may provide biochemical insight into differences between microbial isolates at species and subspecies level, that can provide additional data to genotypic and phenotypic characteristics that can be acquired by WGS and standard molecular methods.

2.8 References

1. Giacinti, G., et al., *Prevalence and characterization of methicillin-resistant Staphylococcus aureus carrying mecA or mecC and methicillin-susceptible Staphylococcus aureus in dairy sheep farms in central Italy.* Journal of dairy science, 2017. **100**(10): p. 7857-7863.
2. Prenafeta, A., et al., *Biofilm production characterization of mecA and mecC methicillin-resistant Staphylococcus aureus isolated from bovine milk in Great Britain.* Journal of dairy science, 2014. **97**(8): p. 4838-4841.
3. Conceição, T., H. de Lencastre, and M. Aires-de-Sousa, *Frequent isolation of methicillin resistant Staphylococcus aureus (MRSA) ST398 among healthy pigs in Portugal.* PloS one, 2017. **12**(4): p. e0175340.
4. Gómez, P., et al., *High prevalence of methicillin-resistant Staphylococcus aureus (MRSA) carrying the mecC gene in a semi-extensive red deer (Cervus elaphus hispanicus) farm in Southern Spain.* Veterinary microbiology, 2015. **177**(3-4): p. 326-331.
5. Paterson, G., et al., *Prevalence and properties of mecC methicillin-resistant Staphylococcus aureus (MRSA) in bovine bulk tank milk in Great Britain.* Journal of Antimicrobial Chemotherapy, 2013. **69**(3): p. 598-602.
6. Oniciuc, E.-A., et al., *Presence of methicillin-resistant Staphylococcus aureus in the food chain.* Trends in Food Science & Technology, 2017. **61**: p. 49-59.
7. CDC, *Antibiotic resistance threats in the United States, 2019*. 2019, U.S. Department of Health and Human Services: Atlanta, GA.
8. Martin, P., et al., *Prevalence of antibiotic-resistant organisms in Canadian Hospitals. Comparison of point-prevalence survey results from 2010, 2012, and 2016.* Infection Control & Hospital Epidemiology, 2019. **40**(1): p. 53-59.
9. Suetens, C., et al., *Prevalence of healthcare-associated infections, estimated incidence and composite antimicrobial resistance index in acute care hospitals and long-term care facilities: results from two European point prevalence surveys, 2016 to 2017.* Eurosurveillance, 2018. **23**(46).
10. Long, S.S., C.G. Prober, and M. Fischer, *Principles and Practice of Pediatric Infectious Diseases E-Book*. 2017: Elsevier Health Sciences.
11. Becker, K., *Pathogenesis of Staphylococcus aureus*, in *Staphylococcus aureus*. 2018, Elsevier. p. 13-38.
12. Organization, W.H., *Antimicrobial resistance: global report on surveillance*. 2014: World Health Organization.
13. Freeman, J. and A. Morris, *Rapid detection of methicillin-resistant Staphylococcus aureus*. 2010, Uptodate. Waltham: UpToDate.
14. Ford, L., et al., *Incorporating Whole-Genome Sequencing into Public Health Surveillance: Lessons from Prospective Sequencing of Salmonella Typhimurium in Australia.* Foodborne pathogens and disease, 2018.
15. Peacock, S.J. and G.K. Paterson, *Mechanisms of methicillin resistance in Staphylococcus aureus.* Annual review of biochemistry, 2015. **84**: p. 577-601.
16. Fang, H. and G. Hedin, *Use of cefoxitin-based selective broth for improved detection of methicillin-resistant Staphylococcus aureus.* Journal of clinical microbiology, 2006. **44**(2): p. 592-594.

17. Fernandes, C.J., L.A. Fernandes, and P. Collignon, *Cefoxitin resistance as a surrogate marker for the detection of methicillin-resistant Staphylococcus aureus.* Journal of Antimicrobial Chemotherapy, 2005. **55**(4): p. 506-510.
18. Feßler, A.T., et al., *Antimicrobial Resistance Properties of Staphylococcus aureus*, in *Staphylococcus aureus*. 2018, Elsevier. p. 57-85.
19. Porrero, M.C., et al., *Detection of mecC-Methicillin-resistant Staphylococcus aureus isolates in river water: a potential role for water in the environmental dissemination.* Environmental microbiology reports, 2014. **6**(6): p. 705-708.
20. Dupieux, C., et al., *Variable performance of four commercial chromogenic media for detection of methicillin-resistant Staphylococcus aureus isolates harbouring mecC.* International journal of antimicrobial agents, 2017. **50**(2): p. 263-265.
21. Stein, C., et al., *Rethinking the Molecular Diagnostics for Methicillin-Resistant Staphylococcus aureus.* Infection Control & Hospital Epidemiology, 2018. **39**(4): p. 495-496.
22. Matouskova, I. and V. Janout, *Current knowledge of methicillin-resistant Staphylococcus aureus and community-associated methicillin-resistant Staphylococcus aureus.* Biomedical Papers, 2008. **152**(2): p. 191-202.
23. Levesque, S., et al., *Molecular epidemiology and antimicrobial susceptibility profiles of methicillin-resistant Staphylococcus aureus blood culture isolates: results of the Quebec Provincial Surveillance Programme.* Epidemiology & Infection, 2015. **143**(7): p. 1511-1518.
24. Bal, A., et al., *Genomic insights into the emergence and spread of international clones of healthcare-, community-and livestock-associated meticillin-resistant Staphylococcus aureus: blurring of the traditional definitions.* Journal of global antimicrobial resistance, 2016. **6**: p. 95-101.
25. Ramos, S., et al., *Enterococci, from harmless bacteria to a pathogen.* Microorganisms, 2020. **8**(8): p. 1118.
26. Kristich, C.J., L.B. Rice, and C.A. Arias, *Enterococcal infection—treatment and antibiotic resistance.* Enterococci: From commensals to leading causes of drug resistant infection [Internet], 2014.
27. Gorrie, C., et al., *Genomics of vancomycin-resistant Enterococcus faecium.* Microbial genomics, 2019. **5**(7).
28. Courvalin, P., *Vancomycin resistance in gram-positive cocci.* Clinical Infectious Diseases, 2006. **42**(Supplement_1): p. S25-S34.
29. Lévesque, S., et al., *A side by side comparison of Bruker Biotyper and VITEK MS: utility of MALDI-TOF MS technology for microorganism identification in a public health reference laboratory.* PloS one, 2015. **10**(12): p. e0144878.
30. Bard, J.D. and F. Lee, *Why can't we just use PCR? The role of genotypic versus phenotypic testing for antimicrobial resistance testing.* Clinical microbiology newsletter, 2018. **40**(11): p. 87-95.
31. Pincus, D.H., *Microbial identification using the bioMérieux Vitek® 2 system.* Encyclopedia of Rapid Microbiological Methods. Bethesda, MD: Parenteral Drug Association, 2006: p. 1-32.
32. Ligozzi, M., et al., *Evaluation of the VITEK 2 system for identification and antimicrobial susceptibility testing of medically relevant gram-positive cocci.* Journal of clinical microbiology, 2002. **40**(5): p. 1681-1686.

33. Stokkou, S., et al., *Direct disk diffusion test using European Clinical Antimicrobial Susceptibility Testing breakpoints provides reliable results compared with the standard method.* European Journal of Microbiology and Immunology, 2015. **5**(1): p. 103-111.
34. Roisin, S., et al., *Evaluation of new Vitek 2 card and disk diffusion method for determining susceptibility of Staphylococcus aureus to oxacillin.* Journal of clinical microbiology, 2008. **46**(8): p. 2525-2528.
35. Singhal, N., et al., *MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis.* Frontiers in microbiology, 2015. **6**: p. 791.
36. Ge, M.-C., et al., *Routine identification of microorganisms by matrix-assisted laser desorption ionization time-of-flight mass spectrometry: success rate, economic analysis, and clinical outcome.* Journal of Microbiology, Immunology and Infection, 2017. **50**(5): p. 662-668.
37. Valentine, N., et al., *Effect of culture conditions on microorganism identification by matrix-assisted laser desorption ionization mass spectrometry.* Applied and Environmental Microbiology, 2005. **71**(1): p. 58-64.
38. Lüthje, P., et al., *Identification of microorganisms grown on chromogenic media by MALDI-TOF MS.* Journal of microbiological methods, 2017. **136**: p. 17-20.
39. Guo, L., et al., *Comparative study of MALDI-TOF MS and VITEK 2 in bacteria identification.* Journal of thoracic disease, 2014. **6**(5): p. 534.
40. TeKippe, E.M. and C.-A. Burnham, *Evaluation of the Bruker Biotyper and VITEK MS MALDI-TOF MS systems for the identification of unusual and/or difficult-to-identify microorganisms isolated from clinical specimens.* European journal of clinical microbiology & infectious diseases, 2014. **33**(12): p. 2163-2171.
41. Lasch, P., et al., *Insufficient discriminatory power of MALDI-TOF mass spectrometry for typing of Enterococcus faecium and Staphylococcus aureus isolates.* Journal of microbiological methods, 2014. **100**: p. 58-69.
42. Patel, R., *MALDI-TOF MS for the diagnosis of infectious diseases.* Clinical chemistry, 2015. **61**(1): p. 100-111.
43. Schubert, S. and M. Kostrzewa, *MALDI-TOF MS in the Microbiology Laboratory: Current Trends.* Current issues in molecular biology, 2017. **23**: p. 17-20.
44. Oviaño, M., et al., *Universal protocol for the rapid automated detection of carbapenem-resistant Gram-negative bacilli directly from blood cultures by matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF/MS).* International journal of antimicrobial agents, 2016. **48**(6): p. 655-660.
45. Mitchell, S. and K. Alby, *Performance of microbial identification by MALDI-TOF MS and susceptibility testing by VITEK 2 from positive blood cultures after minimal incubation on solid media.* European Journal of Clinical Microbiology & Infectious Diseases, 2017. **36**(11): p. 2201-2206.
46. Barberino, M.G., et al., *Direct identification from positive blood broth culture by matrix-assisted laser desorption-ionization time-of-flight mass spectrometry (MALDI-TOF MS).* Brazilian Journal of Infectious Diseases, 2017. **21**(3): p. 339-342.
47. Ghebremedhin, B., et al., *MALDI-TOF MS based carbapenemase detection from culture isolates and from positive blood culture vials.* Annals of clinical microbiology and antimicrobials, 2016. **15**(1): p. 5.

48. Oviaño, M., et al., *Rapid direct detection of carbapenemase-producing Enterobacteriaceae in clinical urine samples by MALDI-TOF MS analysis.* Journal of Antimicrobial Chemotherapy, 2017. **72**(5): p. 1350-1354.
49. Steensels, D., et al., *MALDI-TOF MS typing of a nosocomial methicillin-resistant Staphylococcus aureus outbreak in a neonatal intensive care unit.* Acta Clinica Belgica, 2017. **72**(4): p. 219-225.
50. Shan, W., et al., *Rapid identification of methicillin-resistant staphylococcus aureus (MRSA) by the vitek MS saramis system.* Current microbiology, 2016. **72**(1): p. 29-32.
51. Griffin, P.M., et al., *Use of matrix-assisted laser desorption ionization–time of flight mass spectrometry to identify vancomycin-resistant enterococci and investigate the epidemiology of an outbreak.* Journal of clinical microbiology, 2012. **50**(9): p. 2918-2931.
52. Burckhardt, I. and S. Zimmermann, *Susceptibility testing of bacteria using MALDI-TOF mass spectrometry.* Frontiers in microbiology, 2018. **9**.
53. Kwon, S.S., et al., *Performance of Matrix-Assisted Laser Desorption Ionization Time-of-Fight Mass Spectrometry for Rapid Discrimination of Methicillin-Resistant Staphylococcus aureus (MRSA): First Report of a Relation Between Protein Peaks and MRSA spa Type.* Annals of laboratory medicine, 2017. **37**(6): p. 553-555.
54. Brackmann, M., et al., *Antimicrobial resistance Classification using MALDI-TOF-MS is not that easy: Lessons from vancomycin-resistant Enterococcus faecium.* Clinical Microbiology and Infection, 2019.
55. Hu, Y., et al., *Evaluation of Staphylococcus aureus subtyping module for MRSA detection based on MALDI-TOF MS.* Frontiers in microbiology, 2019. **10**: p. 2504.
56. Timke, M., et al. *Instant detection of PSM positive MRSA in MALDI biotyper workflow*. in *ECCMID Meeting*. 2017.
57. Cordovana, M., et al., *MALDI-TOF bacterial subtyping to detect antibiotic resistance.* Clinical Mass Spectrometry, 2019.
58. Holzknecht, B., et al., *Typing of vancomycin-resistant enterococci with MALDI-TOF mass spectrometry in a nosocomial outbreak setting.* Clinical Microbiology and Infection, 2018. **24**(10): p. 1104. e1-1104. e4.
59. Rödel, J., et al., *Use of MALDI-TOF mass spectrometry to detect nosocomial outbreaks of Serratia marcescens and Citrobacter freundii.* European Journal of Clinical Microbiology & Infectious Diseases, 2019. **38**(3): p. 581-591.
60. Neoh, H.-m., et al., *Pulsed-field gel electrophoresis (PFGE): A review of the "gold standard" for bacteria typing and current alternatives.* Infection, Genetics and Evolution, 2019: p. 103935.
61. Janda, J.M. and S.L. Abbott, *16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls.* Journal of clinical microbiology, 2007. **45**(9): p. 2761-2764.
62. Srinivasan, R., et al., *Use of 16S rRNA gene for identification of a broad range of clinically relevant bacterial pathogens.* PloS one, 2015. **10**(2): p. e0117617.
63. Leach, L., Y. Zhu, and S. Chaturvedi, *Development and validation of a real-time PCR assay for rapid detection of Candida auris from surveillance samples.* Journal of clinical microbiology, 2018. **56**(2): p. e01223-17.

64. Lee, T.D., et al., *Rapid detection of KPC, NDM, and OXA-48-like carbapenemases by real-time PCR from rectal swab surveillance samples.* Journal of clinical microbiology, 2015. **53**(8): p. 2731-2733.
65. Chavda, K.D., et al., *Evaluation of a multiplex PCR assay to rapidly detect Enterobacteriaceae with a broad range of β-lactamases directly from perianal swabs.* Antimicrobial agents and chemotherapy, 2016. **60**(11): p. 6957-6961.
66. Qin, X., et al., *Real-time PCR assay for detection and differentiation of Shiga toxin-producing Escherichia coli from clinical samples.* Journal of clinical microbiology, 2015. **53**(7): p. 2148-2153.
67. Gazin, M., et al., *Evaluation of GeneOhm VanR and Xpert vanA/vanB molecular assays for the rapid detection of vancomycin-resistant enterococci.* European journal of clinical microbiology & infectious diseases, 2012. **31**(3): p. 273-276.
68. Dalpke, A.H., M. Hofko, and S. Zimmermann, *Development of a real-time PCR protocol requiring minimal handling for detection of vancomycin-resistant enterococci with the fully automated BD Max system.* Journal of clinical microbiology, 2016. **54**(9): p. 2321-2329.
69. Espy, M., et al., *Real-time PCR in clinical microbiology: applications for routine laboratory testing.* Clinical microbiology reviews, 2006. **19**(1): p. 165-256.
70. Huletsky, A., et al., *New real-time PCR assay for rapid detection of methicillin-resistant Staphylococcus aureus directly from specimens containing a mixture of staphylococci.* Journal of clinical microbiology, 2004. **42**(5): p. 1875-1884.
71. Hanson, K., et al., *Preclinical assessment of a fully automated multiplex PCR panel for detection of central nervous system pathogens.* Journal of clinical microbiology, 2016. **54**(3): p. 785-787.
72. Perry, J.D., *A decade of development of chromogenic culture media for clinical microbiology in an era of molecular diagnostics.* Clinical microbiology reviews, 2017. **30**(2): p. 449-479.
73. Nonhoff, C., et al., *Comparison of three chromogenic media and enrichment broth media for the detection of methicillin-resistant Staphylococcus aureus from mucocutaneous screening specimens.* European journal of clinical microbiology & infectious diseases, 2009. **28**(4): p. 363-369.
74. Shenoy, E.S., et al., *Impact of rapid screening for discontinuation of methicillin-resistant Staphylococcus aureus contact precautions.* American journal of infection control, 2016. **44**(2): p. 215-221.
75. Mutters, N., et al., *Costs and possible benefits of a two-tier infection control management strategy consisting of active screening for multidrug-resistant organisms and tailored control measures.* Journal of Hospital Infection, 2016. **93**(2): p. 191-196.
76. Garcia, L.S., *Clinical microbiology procedures handbook*. 2010: American Society for Microbiology Press.
77. Miao, J., et al., *Current methodologies on genotyping for nosocomial pathogen methicillin-resistant Staphylococcus aureus (MRSA).* Microbial pathogenesis, 2017. **107**: p. 17-28.
78. Murchan, S., et al., *Harmonization of pulsed-field gel electrophoresis protocols for epidemiological typing of strains of methicillin-resistant Staphylococcus aureus: a single approach developed by consensus in 10 European laboratories and its application for*

tracing the spread of related strains. Journal of clinical microbiology, 2003. **41**(4): p. 1574-1585.
79. Salipante, S.J., et al., *Application of whole-genome sequencing for bacterial strain typing in molecular epidemiology.* Journal of clinical microbiology, 2015. **53**(4): p. 1072-1079.
80. Tenover, F.C., et al., *Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing.* Journal of clinical microbiology, 1995. **33**(9): p. 2233.
81. De Been, M., et al., *Core genome multilocus sequence typing scheme for high-resolution typing of Enterococcus faecium.* Journal of clinical microbiology, 2015. **53**(12): p. 3788-3797.
82. Neumann, B., et al., *A core genome multilocus sequence typing scheme for Enterococcus faecalis.* Journal of clinical microbiology, 2019. **57**(3): p. e01686-18.
83. Inns, T., et al., *Prospective use of whole genome sequencing (WGS) detected a multi-country outbreak of Salmonella Enteritidis.* Epidemiology & Infection, 2017. **145**(2): p. 289-298.
84. Quainoo, S., et al., *Whole-genome sequencing of bacterial pathogens: the future of nosocomial outbreak analysis.* Clinical microbiology reviews, 2017. **30**(4): p. 1015-1063.
85. Vincent, C., et al., *Comparison of advanced whole genome sequence-based methods to distinguish strains of Salmonella enterica serovar Heidelberg involved in foodborne outbreaks in Québec.* Food Microbiology, 2018.
86. Lytsy, B., et al., *Time to review the gold standard for genotyping vancomycin-resistant enterococci in epidemiology: Comparing whole-genome sequencing with PFGE and MLST in three suspected outbreaks in Sweden during 2013–2015.* Infection, Genetics and Evolution, 2017. **54**: p. 74-80.
87. Bortolaia, V., et al., *ResFinder 4.0 for predictions of phenotypes from genotypes.* Journal of Antimicrobial Chemotherapy, 2020. **75**(12): p. 3491-3500.
88. Do Nascimento, V., et al., *Comparison of phenotypic and WGS-derived antimicrobial resistance profiles of enteroaggregative Escherichia coli isolated from cases of diarrhoeal disease in England, 2015–16.* Journal of Antimicrobial Chemotherapy, 2017. **72**(12): p. 3288-3297.
89. Day, M.R., et al., *Comparison of phenotypic and WGS-derived antimicrobial resistance profiles of Salmonella enterica serovars Typhi and Paratyphi.* Journal of Antimicrobial Chemotherapy, 2017.
90. Köser, C.U., et al., *Routine use of microbial whole genome sequencing in diagnostic and public health microbiology.* PLoS pathogens, 2012. **8**(8): p. e1002824.
91. Tetzschner, A.M.M., et al., *In silico genotyping of Escherichia coli isolates for extraintestinal virulence genes by use of whole-genome sequencing data.* Journal of clinical microbiology, 2020. **58**(10).
92. Carattoli, A., et al., *In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing.* Antimicrobial agents and chemotherapy, 2014. **58**(7): p. 3895-3903.
93. Arndt, D., et al., *PHASTER: a better, faster version of the PHAST phage search tool.* Nucleic acids research, 2016. **44**(W1): p. W16-W21.
94. Zhou, Y., et al., *PHAST: a fast phage search tool.* Nucleic acids research, 2011. **39**(suppl_2): p. W347-W352.

95. Raven, K.E., et al., *Complex routes of nosocomial vancomycin-resistant Enterococcus faecium transmission revealed by genome sequencing.* Clinical infectious diseases, 2017. **64**(7): p. 886-893.
96. Brodrick, H.J., et al., *Whole-genome sequencing reveals transmission of vancomycin-resistant Enterococcus faecium in a healthcare network.* Genome medicine, 2016. **8**(1): p. 4.
97. Coll, F., et al., *Longitudinal genomic surveillance of MRSA in the UK reveals transmission patterns in hospitals and the community.* Science translational medicine, 2017. **9**(413): p. eaak9745.
98. Tong, S.Y., et al., *Genome sequencing defines phylogeny and spread of methicillin-resistant Staphylococcus aureus in a high transmission setting.* Genome research, 2015. **25**(1): p. 111-118.
99. Kwong, J.C., et al., *Prospective whole-genome sequencing enhances national surveillance of Listeria monocytogenes.* Journal of clinical microbiology, 2016. **54**(2): p. 333-342.
100. Turner, C.E., et al., *Community outbreaks of group A Streptococcus revealed by genome sequencing.* Scientific reports, 2017. **7**(1): p. 8554.
101. Bertelli, C. and G. Greub, *Rapid bacterial genome sequencing: methods and applications in clinical microbiology.* Clinical Microbiology and Infection, 2013. **19**(9): p. 803-813.
102. Rodriguez-Saona, L., et al., *Rapid detection and identification of bacterial strains by Fourier transform near-infrared spectroscopy.* Journal of agricultural and food chemistry, 2001. **49**(2): p. 574-579.
103. Marques, A.S., et al., *Near-infrared spectroscopy and variable selection techniques to discriminate Pseudomonas aeruginosa strains in clinical samples.* Microchemical Journal, 2016. **124**: p. 306-310.
104. de Sousa Marques, A., et al., *The use of near infrared spectroscopy and multivariate techniques to differentiate Escherichia coli and Salmonella Enteritidis inoculated into pulp juice.* Journal of microbiological methods, 2013. **93**(2): p. 90-94.
105. Kammies, T.-L., et al., *Differentiation of foodborne bacteria using NIR hyperspectral imaging and multivariate data analysis.* Applied microbiology and biotechnology, 2016. **100**(21): p. 9305-9320.
106. Sørensen, L.K. and R. Jepsen, *Detection of cheese batches exposed to Clostridium tyrobutyricum spoilage by near infrared spectroscopy.* Journal of Near Infrared Spectroscopy, 1997. **5**(2): p. 91-97.
107. Suthiluk, P., et al., *Possibility of using near infrared spectroscopy for evaluation of bacterial contamination in shredded cabbage.* International journal of food science & technology, 2008. **43**(1): p. 160-165.
108. Alexandrakis, D., G. Downey, and A.G. Scannell, *Rapid non-destructive detection of spoilage of intact chicken breast muscle using near-infrared and Fourier transform mid-infrared spectroscopy and multivariate statistics.* Food and Bioprocess Technology, 2012. **5**(1): p. 338-347.
109. Lu, X., et al., *Detecting and tracking nosocomial methicillin-resistant Staphylococcus aureus using a microfluidic SERS biosensor.* Analytical chemistry, 2013. **85**(4): p. 2320-2327.
110. Lu, X., et al., *Application of mid-infrared and Raman spectroscopy to the study of bacteria.* Food and Bioprocess Technology, 2011. **4**(6): p. 919-935.

111. Maquelin, K., et al., *Raman spectroscopic method for identification of clinically relevant microorganisms growing on solid culture medium.* Analytical Chemistry, 2000. **72**(1): p. 12-19.
112. Beier, B.D., R.G. Quivey, and A.J. Berger, *Identification of different bacterial species in biofilms using confocal Raman microscopy.* Journal of biomedical optics, 2010. **15**(6): p. 066001.
113. Jarvis, R.M. and R. Goodacre, *Discrimination of bacteria using surface-enhanced Raman spectroscopy.* Analytical Chemistry, 2004. **76**(1): p. 40-47.
114. Kloß, S., et al., *Destruction-free procedure for the isolation of bacteria from sputum samples for Raman spectroscopic analysis.* Analytical and bioanalytical chemistry, 2015. **407**(27): p. 8333-8341.
115. Boardman, A.K., et al., *Rapid detection of bacteria from blood with surface-enhanced Raman spectroscopy.* Analytical chemistry, 2016. **88**(16): p. 8026-8035.
116. Buijtels, P.C.A.M., et al., *Rapid identification of mycobacteria by Raman spectroscopy.* Journal of clinical microbiology, 2008. **46**(3): p. 961-965.
117. Kusters, J., et al., *Raman spectroscopy-based identification of nosocomial outbreaks of the clonal bacterium Escherichia coli.* European Journal of Clinical Microbiology & Infectious Diseases, 2016. **35**(1): p. 83-87.
118. Premasiri, W., et al., *Rapid urinary tract infection diagnostics by surface-enhanced Raman spectroscopy (SERS): identification and antibiotic susceptibilities.* Analytical and bioanalytical chemistry, 2017. **409**(11): p. 3043-3054.
119. Wenning, M. and S. Scherer, *Identification of microorganisms by FTIR spectroscopy: perspectives and limitations of the method.* Applied microbiology and biotechnology, 2013. **97**(16): p. 7111-7120.
120. Alvarez-Ordóñez, A., J. Halisch, and M. Prieto, *Changes in Fourier transform infrared spectra of Salmonella enterica serovars Typhimurium and Enteritidis after adaptation to stressful growth conditions.* International journal of food microbiology, 2010. **142**(1-2): p. 97-105.
121. Baldauf, N.A., et al., *Effect of selective growth media on the differentiation of Salmonella enterica serovars by Fourier-transform mid-infrared spectroscopy.* Journal of microbiological methods, 2007. **68**(1): p. 106-114.
122. Carranza, L., et al., *EVALUATION OF FPA-FTIR SPECTROSCOPY AS A TOOL IN THE DIFFERENTIATION OF CAMPYLOBACTER JEJUNI FROM CAMPYLOBACTER COLI ISOLATED FROM RETAIL CHICKEN SAMPLES.* Journal of Food Safety, 2012. **32**(3): p. 289-295.
123. Silva, S., et al., *Discrimination of clinically relevant Candida species by Fourier-transform infrared spectroscopy with attenuated total reflectance (FTIR-ATR).* RSC Advances, 2016. **6**(94): p. 92065-92072.
124. Moreirinha, C., et al., *MIR spectroscopy as alternative method for further confirmation of foodborne pathogens Salmonella spp. and Listeria monocytogenes.* Journal of food science and technology, 2018. **55**(10): p. 3971-3978.
125. Rodrigues, C., et al., *On the front line of Klebsiella pneumoniae surface structures understanding: establishment of Fourier Transform Infrared (FT-IR) spectroscopy as a capsule typing method.* bioRxiv, 2019: p. 554816.

126. Sousa, C., et al., *Development of a FTIR-ATR based model for typing clinically relevant Acinetobacter baumannii clones belonging to ST98, ST103, ST208 and ST218.* Journal of Photochemistry and Photobiology B: Biology, 2014. **133**: p. 108-114.
127. Campos, J., et al., *Discrimination of non-typhoid Salmonella serogroups and serotypes by Fourier Transform Infrared Spectroscopy: A comprehensive analysis.* International journal of food microbiology, 2018. **285**: p. 34-41.
128. Lam, T.M.L., *Rapid Identification and Classification of Escherichia Coli and Shigella by Attenuated Total Reflectance-Fourier Transform Infrared Spectroscopy*. 2017, McGill University Libraries.
129. Lam, L.M., et al., *Reagent-Free Identification of Clinical Yeasts Using Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy.* Journal of clinical microbiology, 2019: p. JCM. 01739-18.
130. Zarnowiec, P., et al., *Chemometric analysis of attenuated total reflectance infrared spectra of Proteus mirabilis strains with defined structures of LPS.* Innate immunity, 2016. **22**(5): p. 325-335.
131. Davis, R., G. Paoli, and L. Mauer, *Evaluation of Fourier transform infrared (FT-IR) spectroscopy and chemometrics as a rapid approach for sub-typing Escherichia coli O157: H7 isolates.* Food microbiology, 2012. **31**(2): p. 181-190.
132. Wohlmeister, D., et al., *Differentiation of Candida albicans, Candida glabrata, and Candida krusei by FT-IR and chemometrics by CHROMagar™ Candida.* Journal of microbiological methods, 2017. **141**: p. 121-125.
133. Amiali, N.M., et al., *Rapid identification of community-associated methicillin-resistant Staphylococcus aureus by Fourier transform infrared spectroscopy.* Diagnostic microbiology and infectious disease, 2011. **70**(2): p. 157-166.
134. Preisner, O.E., et al., *Discrimination of Salmonella enterica serotypes by Fourier transform infrared spectroscopy.* Food Research International, 2012. **45**(2): p. 1058-1064.
135. Lecellier, A., et al., *Implementation of an FTIR spectral library of 486 filamentous fungi strains for rapid identification of molds.* Food microbiology, 2015. **45**: p. 126-134.
136. Colabella, C., et al., *Merging FT-IR and NGS for simultaneous phenotypic and genotypic identification of pathogenic Candida species.* PloS one, 2017. **12**(12): p. e0188104.
137. Dinkelacker, A.G., et al., *Typing and Species Identification of Clinical Klebsiella Isolates by Fourier Transform Infrared Spectroscopy and Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry.* Journal of clinical microbiology, 2018. **56**(11): p. e00843-18.
138. Martak, D., et al., *Fourier-transform infrared spectroscopy can quickly type gram-negative bacilli responsible for hospital outbreaks.* Frontiers in Microbiology, 2019. **10**: p. 1440.
139. Taha, M., et al., *Use of Fourier transform infrared spectroscopy (FTIR) spectroscopy for rapid and accurate identification of yeasts isolated from human and animals.* International journal of veterinary science and medicine, 2013. **1**(1): p. 15-20.
140. Nagib, S., et al., *Identification of Trueperella pyogenes isolated from bovine mastitis by Fourier transform infrared spectroscopy.* PloS one, 2014. **9**(8): p. e104654.
141. Wenning, M., et al., *Identification and differentiation of food-related bacteria: a comparison of FTIR spectroscopy and MALDI-TOF mass spectrometry.* Journal of microbiological methods, 2014. **103**: p. 44-52.

142. Al-Mariri, A., et al., *IDENTIFICATION OF LISTERIA SPECIES BY FOURIERTRANSFORM INFRARED SPECTROSCOPY.* Bulgarian Journal of Veterinary Medicine, 2019. **22**(3).
143. Bağcıoğlu, M., et al., *Detection and identification of Bacillus cereus, Bacillus cytotoxicus, Bacillus thuringiensis, Bacillus mycoides and Bacillus weihenstephanensis via machine learning based FTIR Spectroscopy.* Frontiers in microbiology, 2019. **10**: p. 902.
144. Dawson, S., et al., *Implementation of Fourier transform infrared spectroscopy for the rapid typing of uropathogenic Escherichia coli.* European journal of clinical microbiology & infectious diseases, 2014. **33**(6): p. 983-988.
145. Sousa, C., et al., *Identification of carbapenem-resistant Acinetobacter baumannii clones using infrared spectroscopy.* Journal of biophotonics, 2014. **7**(5): p. 287-294.
146. Johler, S., et al., *High-resolution subtyping of Staphylococcus aureus strains by means of Fourier-transform infrared spectroscopy.* Systematic and applied microbiology, 2016. **39**(3): p. 189-194.
147. Dieckmann, R., et al., *Rapid characterisation of Klebsiella oxytoca isolates from contaminated liquid hand soap using mass spectrometry, FTIR and Raman spectroscopy.* Faraday discussions, 2016. **187**: p. 353-375.
148. Li, Z., et al., *Rapid subtyping of pathogenic and nonpathogenic Vibrio parahaemolyticus by fourier transform infrared spectroscopy with chemometric analysis.* Journal of microbiological methods, 2018. **155**: p. 70-77.
149. Sharaha, U., et al., *Detection of Extended-Spectrum β-Lactamase-Producing Escherichia coli Using Infrared Microscopy and Machine-Learning Algorithms.* Analytical chemistry, 2019. **91**(3): p. 2525-2530.
150. Lebel, P., et al. *Rapid and Reagent-Free Method for Identification of Enterobacteriaceae Using Attenuated Total Reflectance-Fourier Transform Infrared Spectroscopy*. in *Open Forum Infectious Diseases*. 2016. Oxford University Press.
151. Vallieres, E., et al. *Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy for Rapid Identification of Non-Fermenting Gram-Negative Bacilli Isolated from Patients with Cystic Fibrosis*. in *Open forum infectious diseases*. 2017. Oxford University Press.
152. Ergin, Ç., et al., *ATR-FTIR spectroscopy highlights the problem of distinguishing between Exophiala dermatitidis and E. phaeomuriformis using MALDI-TOF MS.* Microbial ecology, 2016. **71**(2): p. 339-346.
153. Costa, F.S., et al., *Attenuated total reflection Fourier transform-infrared (ATR-FTIR) spectroscopy as a new technology for discrimination between Cryptococcus neoformans and Cryptococcus gattii.* Analytical Methods, 2016. **8**(39): p. 7107-7115.
154. Staniszewska-Slezak, E., et al., *Transmission versus transflection mode in FTIR analysis of blood plasma: is the electric field standing wave effect the only reason for observed spectral distortions?* Analyst, 2015. **140**(7): p. 2412-2421.
155. DeVetter, B.M., et al., *Characterization of the structure of low-e substrates and consequences for IR transflection measurements.* Vibrational Spectroscopy, 2017. **91**: p. 119-127.
156. Filik, J., et al., *Electric field standing wave artefacts in FTIR micro-spectroscopy of biological materials.* Analyst, 2012. **137**(4): p. 853-861.
157. !!! INVALID CITATION !!! {}.

158. Lee, J., *On the non-existence of the so-called "electric field standing wave effect" in transflection FTIR spectra.* Vibrational Spectroscopy, 2017. **90**: p. 104-111.
159. Bassan, P., et al., *The inherent problem of transflection-mode infrared spectroscopic microscopy and the ramifications for biomedical single point and imaging applications.* Analyst, 2013. **138**(1): p. 144-157.
160. Baker, M.J., et al., *Using Fourier transform IR spectroscopy to analyze biological materials.* Nature protocols, 2014. **9**(8): p. 1771.
161. Kochan, K., et al., *Comparison of FTIR transmission and transfection substrates for canine liver cancer detection.* Analyst, 2015. **140**(7): p. 2402-2411.
162. DeSouza, L., T. Kamat, and C. Naik, *Diffuse reflectance infrared fourier transform spectroscopic (DRIFTS) investigation of E. coli, Staphylococcus aureus and Candida albicans.* 2009.
163. Sedman, J., et al., *Infrared Imaging: principles and practices.* Handbook of Vibrational Spectroscopy, 2010.
164. Kirkwood, J., et al., *Differentiation of group I and group II strains of Clostridium botulinum by focal plane array Fourier transform infrared spectroscopy.* Journal of food protection, 2006. **69**(10): p. 2377-2383.
165. Aboualizadeh, E., et al., *Understanding the antimicrobial activity of selected disinfectants against methicillin-resistant Staphylococcus aureus (MRSA).* PloS one, 2017. **12**(10): p. e0186375.
166. Lorenz, B., et al., *Cultivation-free Raman spectroscopic investigations of bacteria.* Trends in microbiology, 2017. **25**(5): p. 413-424.
167. Tang, J.-W., et al., *Comparative analysis of machine learning algorithms on surface enhanced Raman spectra of clinical Staphylococcus species.* Frontiers in Microbiology, 2021. **12**.
168. Yan, S., et al., *Raman spectroscopy combined with machine learning for rapid detection of food-borne pathogens at the single-cell level.* Talanta, 2021. **226**: p. 122195.
169. Wu, C.L., et al., *Proton high-resolution magic angle spinning NMR analysis of fresh and previously frozen tissue of human prostate.* Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, 2003. **50**(6): p. 1307-1311.
170. Ramakrishnan, V. and D.L. Luthria, *Recent applications of NMR in food and dietary studies.* Journal of the Science of Food and Agriculture, 2017. **97**(1): p. 33-42.
171. Zailer, E., Y.B. Monakhova, and B.W. Diehl, *31P NMR Method for Phospholipid Analysis in Krill Oil: Proficiency Testing—A Step toward Becoming an Official Method.* Journal of the American Oil Chemists' Society, 2018. **95**(12): p. 1467-1474.
172. Kang, C., *Applications of in-cell NMR in structural biology and drug discovery.* International journal of molecular sciences, 2019. **20**(1): p. 139.
173. Mazzei, P. and A. Piccolo, *HRMAS NMR spectroscopy applications in agriculture.* Chemical and Biological Technologies in Agriculture, 2017. **4**(1): p. 1-13.
174. Balsgart, N.M., et al., *High throughput identification and quantification of phospholipids in complex mixtures.* Analytical chemistry, 2016. **88**(4): p. 2170-2176.
175. Ahmmed, M.K., et al., *Simple and Efficient One-Pot Extraction Method for Phospholipidomic Profiling of Total Oil and Lecithin by Phosphorus-31 Nuclear Magnetic Resonance Measurements.* Journal of Agricultural and Food Chemistry, 2020. **68**(48): p. 14286-14296.

176. Klinowski, J., *Magic-angle-spinning NMR.* Solid State Ionics, 1985. **16**: p. 3-14.
177. Mobarhan, Y.L., et al., *Effective combined water and sideband suppression for low-speed tissue and in vivo MAS NMR.* Analytical and bioanalytical chemistry, 2017. **409**(21): p. 5043-5055.
178. Li, W., et al., *Methods for acquisition and assignment of multidimensional high-resolution magic angle spinning NMR of whole cell bacteria.* Analytical chemistry, 2005. **77**(18): p. 5785-5792.
179. Li, W., *Multidimensional HRMAS NMR: a platform for in vivo studies using intact bacterial cells.* Analyst, 2006. **131**(7): p. 777-781.
180. Righi, V., et al., *Effects of a small, volatile bacterial molecule on Pseudomonas aeruginosa bacteria using whole cell high-resolution magic angle spinning nuclear magnetic resonance spectroscopy and genomics.* International journal of molecular medicine, 2018. **42**(4): p. 2129-2136.
181. Himmelreich, U., T.C. Sorrell, and H.-M. Daniel, *Nuclear magnetic resonance spectroscopy-based identification of yeast*, in *Human Fungal Pathogen Identification*. 2017, Springer. p. 289-304.
182. Bourne, R., et al., *Identification of Enterococcus, Streptococcus, andStaphylococcus by Multivariate Analysis of Proton Magnetic Resonance Spectroscopic Data from Plate Cultures.* Journal of clinical microbiology, 2001. **39**(8): p. 2916-2923.
183. Himmelreich, U., et al., *A rapid screening test to distinguish between Candida albicans and Candida dubliniensis using NMR spectroscopy.* FEMS microbiology letters, 2005. **251**(2): p. 327-332.
184. Himmelreich, U., et al., *Rapid identification of Candida species by using nuclear magnetic resonance spectroscopy and a statistical classification strategy.* Appl. Environ. Microbiol., 2003. **69**(8): p. 4566-4574.
185. Palama, T.L., et al., *Identification of bacterial species by untargeted NMR spectroscopy of the exo-metabolome.* Analyst, 2016. **141**(15): p. 4558-4561.
186. Rautureau, G.J., et al., *Discrimination of Escherichia coli and Shigella spp. by nuclear magnetic resonance based metabolomic characterization of culture media.* ACS infectious diseases, 2019. **5**(11): p. 1879-1886.
187. Mylonakis, E., et al., *T2 magnetic resonance assay for the rapid diagnosis of candidemia in whole blood: a clinical trial.* Clinical Infectious Diseases, 2015. **60**(6): p. 892-899.
188. Neely, L.A., et al., *T2 magnetic resonance enables nanoparticle-mediated rapid detection of candidemia in whole blood.* Science translational medicine, 2013. **5**(182): p. 182ra54-182ra54.
189. De Angelis, G., et al., *T2Bacteria magnetic resonance assay for the rapid detection of ESKAPEc pathogens directly in whole blood.* Journal of Antimicrobial Chemotherapy, 2018. **73**(suppl_4): p. iv20-iv26.
190. Szymanski, C.M., et al., *Detection of conserved N-linked glycans and phase variable lipo-oligosaccharides and capsules from Campylobacter cells by mass spectrometry and high resolution magic angle spinning NMR spectroscopy.* Journal of Biological Chemistry, 2003.
191. Lee, R.E., et al., *Rapid structural characterization of the arabinogalactan and lipoarabinomannan in live mycobacterial cells using 2D and 3D HR-MAS NMR: structural changes in the arabinan due to ethambutol treatment and gene mutation are observed.* Glycobiology, 2004. **15**(2): p. 139-151.

192. Romaniuk, J.A. and L. Cegelski, *Bacterial cell wall composition and the influence of antibiotics by cell-wall and whole-cell NMR.* Phil. Trans. R. Soc. B, 2015. **370**(1679): p. 20150024.
193. Romaniuk, J.A. and L. Cegelski, *Peptidoglycan and Teichoic Acid Levels and Alterations in S. aureus by Cell-Wall and Whole-Cell NMR.* Biochemistry, 2018.
194. Jachymek, W., et al., *Structures of the O-specific polysaccharides from Yokenella regensburgei (Koserella trabulsii) strains PCM 2476, 2477, 2478, and 2494: high-resolution magic-angle spinning NMR investigation of the O-specific polysaccharides in native lipopolysaccharides and directly on the surface of living bacteria.* Biochemistry, 1999. **38**(36): p. 11788-11795.
195. Gudlavalleti, S.K., et al., *In vivo determination of Neisseria meningitidis serogroup A capsular polysaccharide by whole cell high-resolution magic angle spinning NMR spectroscopy.* Carbohydrate research, 2006. **341**(4): p. 557-562.
196. Wong, A., C. Boutin, and P.M. Aguiar, *1H high resolution magic-angle coil spinning (HR-MACS) μNMR metabolic profiling of whole Saccharomyces cervisiae cells: a demonstrative study.* Frontiers in chemistry, 2014. **2**: p. 38.
197. Zhou, X. and L. Cegelski, *Nutrient-dependent structural changes in S. aureus peptidoglycan revealed by solid-state NMR spectroscopy.* Biochemistry, 2012. **51**(41): p. 8143-8153.
198. Maes, E., et al., *Molecular phenotyping of mannosyltransferases-deficient Candida albicans cells by high-resolution magic angle spinning NMR.* Journal of biochemistry, 2009. **145**(4): p. 413-419.
199. Righi, V., et al., *Live-cell high resolution magic angle spinning magnetic resonance spectroscopy for in vivo analysis of Pseudomonas aeruginosa metabolomics.* Biomedical reports, 2013. **1**(5): p. 707-712.
200. Wilson Jr, G.E., G.S. Jacob, and J. Schaefer, *Solid-state 15N NMR studies of the effects of penicillin on cell-wall metabolism of Aerococcusviridans (Gaffkyahomari).* Biochemical and biophysical research communications, 1985. **126**(3): p. 1006-1012.
201. Palomino-Schätzlein, M., et al., *Optimised protocols for the metabolic profiling of S. cerevisiae by 1 H-NMR and HRMAS spectroscopy.* Analytical and bioanalytical chemistry, 2013. **405**(26): p. 8431-8441.
202. Bundy, J.G., et al., *Discrimination of pathogenic clinical isolates and laboratory strains of Bacillus cereus by NMR-based metabolomic profiling.* FEMS microbiology letters, 2005. **242**(1): p. 127-136.

Chapter 3. Rapid Identification of Clinical Pathogens by Transflection-FTIR Spectroscopy.

3.1 Abstract.

FTIR spectroscopy is gaining interest as a technique for microbial identification and discrimination. Transmission and attenuated total reflectance (ATR) spectral acquisition methods are commonly used, whereas transflection has rarely been used in the past for this application. In this study, isolates were subjected to spectral acquisition by all three acquisition modes to compare and determine whether transflection mode has the capabilities to acquire high quality spectra suitable for microbial analyses. Transflection FTIR spectra showed comparable signal-to-noise levels, and spectral discrimination capabilities, with successful discrimination of *S. aureus* from coagulase negative *Staphylococcus* species. To further assess the capabilities of transflection FTIR spectroscopy for microbial identification, a multi-tiered spectral database was developed, using principal component analysis and support vector machine. The database was constructed using 1095 isolates collected at 4 different microbiology laboratories across Canada, representing 22 genera [14 Gram-negative (GN), 7 Gram-positive (GP) and 1 yeast genera]. At the species level, the database was constructed for the identification of 9 *Staphylococcus* spp., 3 *Enterococcus* spp., and 2 *Shigella* spp., as well as the identification of *L. monocytogenes*. The database was evaluated using two sets of isolates that were independent of isolates used for spectral database construction. The validation set (n = 740) only comprised of genera and species that were represented in the spectral database, while the test set (n = 1793) also included isolates that were not represented in the database. Isolates of non-represented genera and species were used to evaluate the database's ability to provide a no-identification result for isolates with unfamiliar spectral features. The spectral database's identification capabilities were evaluated against reference methods such as VITEK 2, MALDI-TOF MS, and whole genome sequencing. Genus-level identification for GP bacteria achieved 99.6 % and 96.5 % concordance to reference methods for validation and test set respectively, while GN genera achieved 82.8 % and 64.7 % correct identification for validation and test set respectively. Well-represented species in the spectral database such as *E. faecium, E. faecalis,* and *S. aureus* performed well with an overall concordance rate of over 98 % in both validation and test sets, demonstrating successful species-level identification. Genera and species that were lacking representation in the spectral database was unable to make robust correct

identification with high confidence levels. Misidentification of non-represented isolates that were tested also occurred at 51.4 % (n = 36), 81 % (n = 17) and 37.5 % (n = 6) respectively for GN genera, GP genera and staphylococcal species. Further investigation on spectral discrimination between genera and species that are difficult to differentiate are required prior to their addition for database expansion. The study demonstrated that the constructed transflection spectral database demonstrated successful identification capabilities for species that are well represented.

3.2 Introduction

Rapid microbial identification is critical for determining the causative agent of infections, and for providing optimal treatment to patients. There are many identification methods that are currently accepted and used for diagnostic purposes, such as matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) and biochemical based assays. MALDI-TOF MS has become increasingly popular in the past few years, due to its ease in sample preparation and relatively fast turnaround time to results (up to an hour from acquiring isolated colonies on agar) [1]. The disadvantage of MALDI-TOF MS is the high principle capital cost, which prevents laboratories with low budgets from purchase and use [2]. Furthermore, some genera, and species are difficult to identify using these reference methods, such as discrimination between *Shigella* species and *Escherichia coli* and differentiation between *S. pneumoniae* and *S. mitis/oralis* group by MALDI-TOF MS [3, 4]. Automated biochemical-based assay systems, such as VITEK 2 (bioMérieux SA, Marcy l'Étoile, France) and Pheonix (BD, USA) take longer to obtain results, sometimes requiring overnight testing [1].

3.2.1 FTIR spectroscopy in microbial identification/discrimination

Microbial discrimination using infrared spectroscopy at the species and sub-species level has been studied over the past few decades exploring the advantages and limitations for microbial diagnostic purposes [5, 6]. There are three different spectral acquisition modes in FTIR spectroscopy, transmission, ATR and transflection (In studies conducted over the years, various mathematical and multivariate statistical analysis methods have been applied to optimize and evaluate spectral discrimination capabilities of FTIR spectral data. Successful spectral discrimination have been demonstrated using principal component analyses (PCA)[20], hierarchical cluster analysis (HCA)[7], partial least-squares-discriminant analysis (PLS-DA)[21], artificial neural network (ANN)[22, 23], and other machine learning techniques such as support vector machine (SVM)[12], self-organizing map (SOM) and K-nearest neighbour (KNN)[24], or combinations thereof [25, 26].

Although the advantages and limitations of using different spectral acquisition modes were briefly explained by Wenning &Scheer (2013), there has yet to be reports that directly compare the three techniques using the same samples, to minimize variability from sample size, sample preparation, spectral analysis methods [5]. Therefore in this study, a limited set of bacteria was

subjected to spectral acquisition by transmission (FPA-FTIR imaging), ATR and transflection modes to compare the spectral quality, as well as the differences in sample preparation and ease of spectral acquisition. Additionally, due to the lack of evaluation in spectral discrimination and identification of microorganisms using transflection-FTIR spectroscopy, the second portion of the chapter focused on the development, and evaluation of a transflection FTIR spectral database for a limited set of genera and species, mainly focusing on Gram-positive bacteria. Microbial identification results produced by transflection-FTIR spectral database were compared against clinically accepted identification methods, including biochemical-based assays (VITEK 2), MALDI-TOF MS, 16S rRNA sequencing and whole genome sequencing (WGS).

Figure 3.1). Transmission has been the most commonly used spectral acquisition mode for studying microbial discrimination, followed by attenuated total reflectance (ATR). The sampling techniques provide different advantages and disadvantages for use in microbial identification. Additionally, with the use of a light microscope and a focal-plane-array detector, FTIR spectra and images can be acquired from a smaller surface area than what can be achieved using a single-element detector spectrometer. Previous work from the McGill IR group showcased successful discrimination of bacteria collected by FPA-FTIR imaging in transmission mode at genus, species and sub-species level through optimization and standardization of sample preparation, spectral acquisition and spectral processing [7-9].

3.2.1.1 Transmission FTIR spectroscopy

For spectral acquisition in transmission mode, microbial sample is deposited on an IR-transparent substrate such as zinc selenide (ZnSe), silicon (Si) or calcium fluoride (CaF_2). The infrared light passes through the sample on the substrate, and reaches the detector. To minimize artefacts relating to sample thickness heterogeneity, it is important that the samples are deposited as evenly as possible at the optimal thickness (2-8 μm). Many studies have implemented a sample preparation protocol that involves suspending microbial colonies in water and pipetting the sample solution to achieve homogeneity across the surface area [5, 10]. The IR-transparent substrate used for spectral acquisition is not cheap depending on the material chosen (such as ZnSe), and hence would be costly for single-use applications. Proper sterilization would be required to re-use these substrates, to prevent cross-contamination between samples, and to ensure the safety of the laboratory technicians who handle the samples. These IR transparent optical materials also have

different spectral ranges for IR signal, and vary in water solubility, which should be considered when choosing the substrate [10, 11].

3.2.1.2 Attenuated total reflectance (ATR) FTIR spectroscopy

ATR-FTIR spectroscopy utilizes a IR-transparent crystal made of high-refractive index material, such as diamond, zinc selenide or germanium, with a defined angle, where the sample and infrared light interacts. The angle of the crystal should be larger than the critical angle for the infrared beam to achieve total internal reflection, and reach the detector. When the infrared beam reflects off of the ATR-crystal, which is in contact with the analyte, an evanescent wave is attenuated by the sample. The advantage of the ATR mode is that the technique is immune to sample thickness, as long as the sample is in sufficient contact with the ATR-crystal, since the evanescent wave is only capable of penetrating a few microns (wavelength dependent) into the sample [5]. Many studies that utilize ATR spectral acquisition modes for studying bacteria require a lengthy step of air-drying the colony that have been directly deposited on the ATR-crystal. This drying process makes ATR spectral acquisition a very time-consuming method, as each spectral acquisition requires approximately 30 minutes [12]. ATR-FTIR spectra were acquired from wet colonies in protocols developed by the McGill IR Group. The removal of the drying step required by other groups improves the overall spectral acquisition time. Due to the strong absorbance signal from free- and bound-water, the signals from other biochemical constituents are rather limited, in comparison to spectra acquired from dried samples [13]. However, the development and performance of an extensive spectral database for clinical yeast identification was not hindered, and achieved highly accurate identification results relative to MALDI-TOF MS [13].

3.2.1.3 Transflection FTIR spectroscopy

Compared to transmission and ATR spectral acquisition, there are only a handful of studies conducted using reflectance (for both diffuse and specular) for microbial discrimination evaluation [14-17]. Transflection-FTIR spectroscopy's lack of popularity in microbial studies may be due to reported spectral distortions caused by light scattering and electron free-stranding waves when IR light interacts with the sample and the IR reflective surface [18]. However, with application of mathematical pre-processing to the spectra, transflection FTIR spectra can provide the same analytical conclusion in application towards biological materials, such as spectral discrimination between control (healthy) and malignant cells, or discrimination of microorganisms [10, 11].

Transflection mode has a few advantages, first being that the substrates used for sample deposition are cheaper than those used in transmission FTIR spectroscopy [10]. Low-emissivity glass (E-glass) from Kevley technologies (Chesterland, Ohio, USA) provides microscope glass slides with IR reflective coatings. E-glass and other surfaces coated with metallic substances such as gold, silver or zinc oxide, are also cheaper alternative to IR transparent materials. By taking advantage of cheaper materials, IR-reflective substrates can be used for single-use, and easily disposed in microbiology laboratories making it easier to ensure sanitary and safe working area [18]. By using the E-glass, which is optically transparent, samples can be observed under a microscope using visible light, making it available for visual inspection. E-glass has been a popular choice for some time in acquiring FTIR images from cells and tissues for this very reason [10]. Compared to transmission mode, the signal obtained from transflection mode on the same sample would effectively double, since the infrared light passes through the sample twice prior to reaching the detector [10].

3.2.1.4 FPA-FTIR imaging spectroscopy

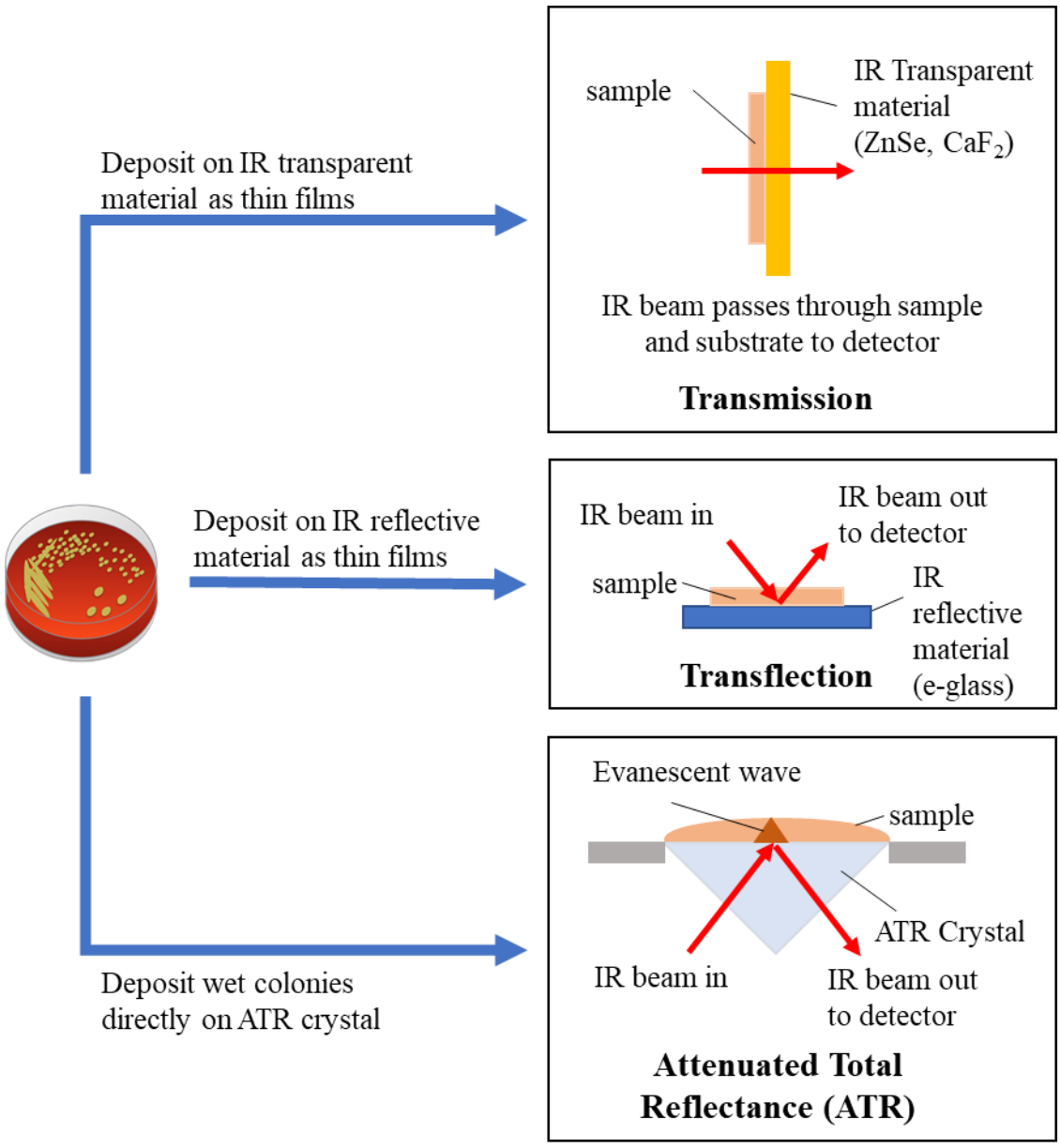
A focal plane array (FPA) -FTIR spectrometer is typically used in IR microspectroscopy, by combining a spectrometer with a microscope. The FPA detector has the ability to acquire tens and thousands of spectra in one spectral acquisition, due to the presence of many elemental detectors combined into a grid form (i.e. 16x16, 64x64 or even 128x128), producing FTIR images with up to 16,384 spectra per image. The FPA detector produces FTIR images with spatial resolutions as high as ~5 microns across an area of 0.3 mm x 0.3 mm in IR microspectrometery. In order to achieve high signal to noise ratio (SNR) spectra in the FTIR images, mercury cadmium telluride (MCT) detector is used. The MCT detector requires the use of liquid nitrogen, to improve the SNR of the spectra acquired. FTIR imaging is useful in studying cells, tissues and any biofluids, as it allows for observation of chemical distribution and interaction across the sample, rather than obtaining one representative spectrum achieved by a single detector spectrometer. All three spectral acquisition modes can be used in FPA-FTIR imaging. In microbial identification applications, FPA-FTIR spectroscopy holds potential in identification from direct clinical samples, or simultaneous identification of multiple organisms present in mixed cultures [19].

3.2.1.5 FTIR spectral processing and data analysis

In studies conducted over the years, various mathematical and multivariate statistical analysis methods have been applied to optimize and evaluate spectral discrimination capabilities of FTIR spectral data. Successful spectral discrimination have been demonstrated using principal component analyses (PCA)[20], hierarchical cluster analysis (HCA)[7], partial least-squares-discriminant analysis (PLS-DA)[21], artificial neural network (ANN)[22, 23], and other machine learning techniques such as support vector machine (SVM)[12], self-organizing map (SOM) and K-nearest neighbour (KNN)[24], or combinations thereof [25, 26].

Although the advantages and limitations of using different spectral acquisition modes were briefly explained by Wenning &Scheer (2013), there has yet to be reports that directly compare the three techniques using the same samples, to minimize variability from sample size, sample preparation, spectral analysis methods [5]. Therefore in this study, a limited set of bacteria was subjected to spectral acquisition by transmission (FPA-FTIR imaging), ATR and transflection modes to compare the spectral quality, as well as the differences in sample preparation and ease of spectral acquisition. Additionally, due to the lack of evaluation in spectral discrimination and identification of microorganisms using transflection-FTIR spectroscopy, the second portion of the chapter focused on the development, and evaluation of a transflection FTIR spectral database for a limited set of genera and species, mainly focusing on Gram-positive bacteria. Microbial identification results produced by transflection-FTIR spectral database were compared against clinically accepted identification methods, including biochemical-based assays (VITEK 2), MALDI-TOF MS, 16S rRNA sequencing and whole genome sequencing (WGS).

Figure 3.1 Description of the three FTIR spectral acquisition modes



3.3 Materials and Methods

3.3.1 Isolates for construction of transflection-FTIR spectral database of select clinically relevant genera and species

1835 bacterial and yeast isolates (6692 spectra), representing 22 genera were employed in the development and evaluation of the spectral database (Table 3.1 and Table 3.2). 60 % of isolates (1095 isolates, 4020 spectra) were used for database construction, and the remaining 40 % (740 isolates, 2672 spectra) were used for the evaluation of the database's ability to correctly identify isolates. Isolates were provided by microbiology laboratories at the McGill University Health Centre (MUHC), Centre hospitalier universitaire Sainte Justine (CHUSJ), Laboratoire de santé publique du Québec (LSPQ) and Canadian Food Inspection Agency (CFIA) where they have been previously identified using reference methods such as MALDI-TOF MS, VITEK 2 and whole genome sequencing (WGS) and stored in 10 % glycerol stock at -80 °C. Isolates were thawed, cultured onto Columbia agar with 5 % sheep blood (BAP), and incubated at 35 °C for 24 hours. Prior to spectral acquisition, samples were sub-cultured and incubated following the same procedure.

3.3.2 Isolates for evaluation of the developed transflection-FTIR spectral database

1793 additional clinical isolates (from skin, blood, urine, and other specimens) were collected retrospectively and prospectively at hospitals, research institutes and reference laboratories (Table 3.2) as part of a second evaluation of the developed spectral database. Isolates were identified at the respective locations, using one or a combination of reference methods, such as MALDI-TOF MS (VITEK MS, bioMérieux SA, Marcy l'Étoile, France; MALDI Biotyper, Bruker Corporation, Billerica, MA, USA), VITEK 2 (VITEK MS, bioMérieux SA, Marcy l'Étoile, France), 16S rRNA sequencing and WGS. These samples were collected independently from those used to develop the spectral database.

3.3.3 FTIR spectral acquisition and pre-processing parameters

Isolated colonies were picked and deposited onto MirrIR IR reflective slide (E-glass) (Kevley Technologies, Chesterland, OH) using a sterile, disposable 1µl loop, and air dried into a thin film. Transflection-FTIR spectra were acquired using a Cary 630 FTIR spectrometer (Agilent Technologies, Santa Clara, CA), with a 10° specular reflectance accessory (Agilent Technologies,

Santa Clara, CA), or a SurveyIR microscope (Czitek, Danbury, CT, USA) with 1 mm aperture. Spectra were acquired with 64 coadded scans in the spectral range 650 to 4000 cm^{-1} with 8 cm^{-1} resolution and a zero-filling factor of 8. A background spectrum was collected from a bare surface of the E-glass prior to every spectral acquisition from samples, with the same spectral acquisition parameter settings. Triplicate spectra were acquired per isolate, from different surface areas of the sample deposition(s). Reproducibility of the triplicate spectra acquired for each isolate was examined for spectral quality check, based on proximity in a dendrogram produced by HCA of the FTIR spectra based on spectral similarity [13]. Spectra were processed using first derivative and vector normalization in order to remove variability caused by spectral baseline shifts and sample thickness respectively, prior to employing multivariate statistical analyses for discrimination and identification.

3.3.4 Comparison of spectral acquisition modes

Isolates collected in Brisbane, Australia were used to compare the three spectral acquisition modes, transmission, ATR and transflection, based on their ease of instrument use and sample preparation, as well as spectral quality and spectral discriminatory capabilities. ATR-FTIR spectra were acquired using a single-bounce diamond ATR accessory on the same FTIR spectrometer as the one used for transflection-FTIR spectral acquisition. Triplicate spectra were acquired per sample in the spectral range 4000 cm^{-1} to 650 cm^{-1} with 64 co-added scans, and zero-filling factor of 2 was employed. A background spectrum was acquired with the same parameters from a bare ATR crystal. Transmission FPA-FTIR spectroscopy acquired 16,384 spectra (FTIR image comprised of 128 x 128 pixels) simultaneously from each isolate per image. Triplicate images were acquired per isolate on ZnSe and CaF_2 substrate. For FPA images, each image was filtered to remove spectral data which had inadequate signal to noise ratio, and insufficient absorbance in the amide I and II region. All images were processed using an in-house software, to retrieve high-quality spectra from each image (Cognisolve Inc, Montreal, QC, Canada). Like the transflection FTIR spectra, ATR- and transmission FTIR spectra were processed by taking the first derivative and vector normalized to remove the effect of baseline shifts and sample thickness variability respectively.

3.3.5 Transflection-FTIR spectral database construction

Upon spectral pre-processing and removal of spectra that did not meet minimum spectral quality requirements, spectral database was constructed in a hierarchical fashion, where isolates were categorized based on their Gram morphology, followed by genera and species discrimination (Figure 3.2) [13]. Variable reduction was conducted by employing PCA to the spectral region 1350-800 cm^{-1}. Sixteen principal components (PC) were kept and used to create the SVM-based spectral database. Each PC score that was used for subsequent analyses explained > 1 % of total variance between spectral data, resulting in a total explained variables of 87.4 % by the 16 PCs. Using radial basis function support vector machine (SVM) in JMP Pro ver. 15.2.0 (SAS Institute Inc., Cary, NC, USA) prediction models were developed at each level of the database. At each level of identification, prediction models were developed in binary and multiclass PCA-SVM analyses. The SVM variables, cost and gamma, were maintained at cost = 1 and gamma = 0.0625, throughout all levels of the spectral database. Gamma value was calculated based on the number of datapoints, where gamma = $\frac{1}{number\ of\ datapoints\ used\ for\ SVM\ calculation}$. Select genera and species were represented in the database (Table 3.2). The database was then used to predict newly acquired spectral data to determine whether these isolates could be correctly identified at the genus and/or species level. Due to the limited genera and species representation in the database, the database was also tested whether or not it was able to identify non-represented species and produce results as “non-identifiable”.

Table 3.1 Distribution of isolates between database construction (training), validation and test of transflection-FTIR spectral database.

Location	Time Period	Sample Type	Isolates			Reference ID Method(s)
			Training	Validation	Test	
MUHC (Montreal, Canada)	2017-2019	Clinical	570	382	1025	MALDI-TOF MS
NML (Winnipeg, Canada)	May 2018	Clinical	0	0	172	spa typing and PCR
CHU-Sainte-Justine (Montreal, Canada)	July-November 2018	Clinical	274	174	533	VITEK 2
QIMR Berghofer (Brisbane, Australia)	August-November 2018	Clinical	0	0	63	MALDI-TOF MS
CFIA-Ottawa (Ottawa, Canada)	May 2019	Food/Feed-borne	164	117	0	WGS
McGill Macdonald Campus (Sainte-Anne-de-Bellevue, Canada)	2019-2020	Cow Mastitis	87	67	0	MALDI-TOF MS
TOTAL			**1095**	**740**	**1793**	

Figure 3.2 The overview of genera and species represented in the tier-wise transflection-FTIR spectral database

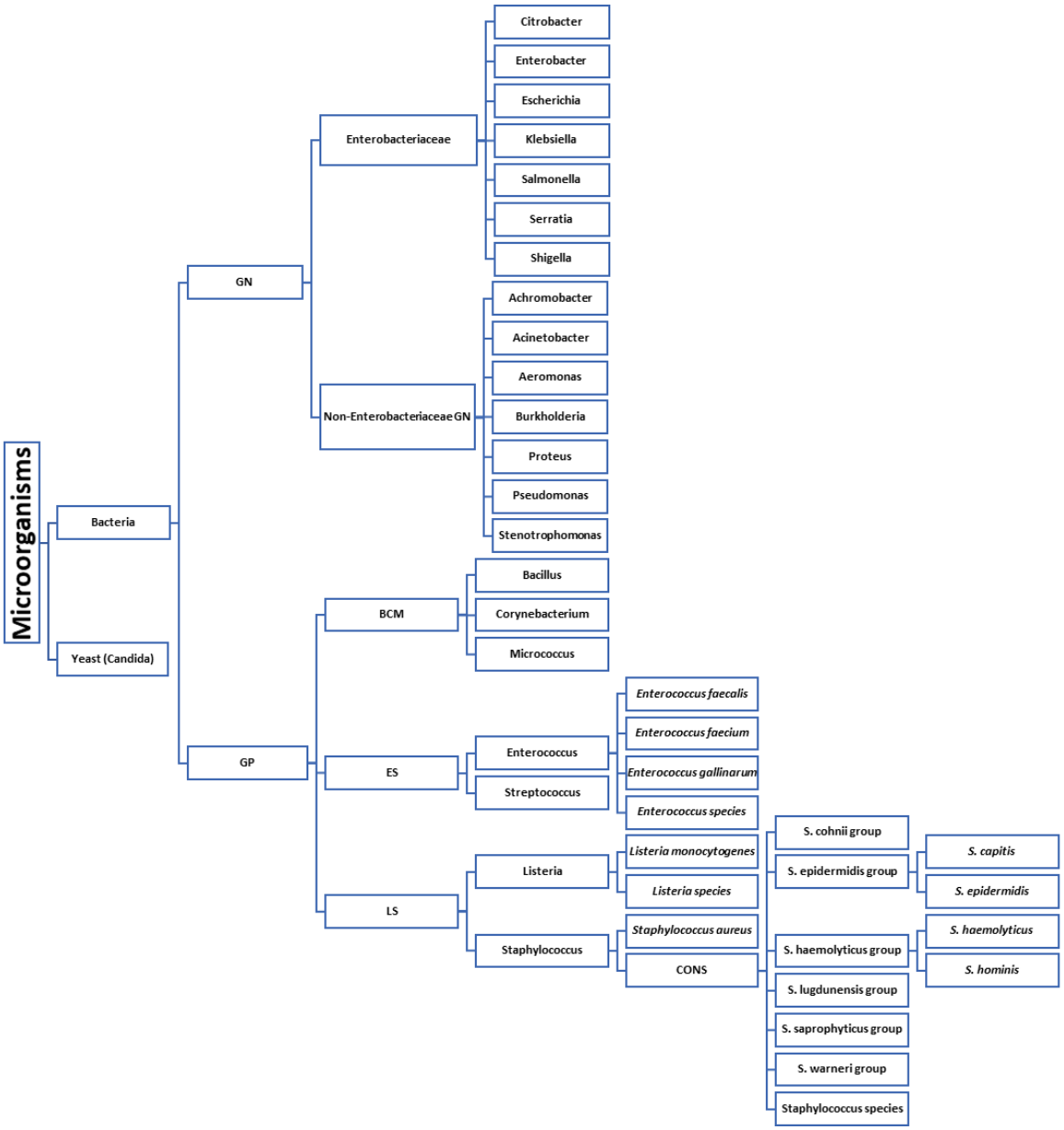


Table 3.2 Number of isolates and spectra used to represent select genera and species in the transflection FTIR spectral database.

Genus	Species	Isolates in Database	Spectra in Database
Achromobacter	*species*	22	74
Acinetobacter	*species*	8	24
Aeromonas	*species*	14	53
Burkholderia	*species*	13	42
Citrobacter	*amalonaticus*	2	6
Citrobacter	*freundii*	10	48
Citrobacter	*koseri*	3	18
Citrobacter	*youngae*	2	6
Citrobacter	*species*	4	12
Enterobacter	*cloacae-complex*	17	76
Enterobacter	*species*	5	15
Escherichia	*coli*	59	201
Klebsiella	*aerogenes*	9	51
Klebsiella	*oxytoca*	10s	49
Klebsiella	*pneumoniae*	22	80
Proteus	*species*	11	33
Pseudomonas	*aeruginosa*	33	114
Salmonella	*bongori*	3	9
Salmonella	*enterica*	94	316
Serratia	*marcescens*	8	36
Shigella	*flexneri*	14	42
Shigella	*sonnei*	11	48
Stenotrophomonas	*species*	15	49
Vibrio	*species*	4	12
Bacillus	*cereus*	7	21
Bacillus	*megaterium*	4	12
Bacillus	*subtilis*	1	3
Bacillus	*species*	4	10
Corynebacterium	*amylolactum*	1	3
Corynebacterium	*aurimucosum*	1	3
Corynebacterium	*bovis*	1	3
Corynebacterium	*imitans*	1	3
Corynebacterium	*jeikeium*	1	3
Corynebacterium	*striatum*	2	6

(continued)

Genus	Species	Isolates in Database	Spectra in Database
Corynebacterium	*tuberculosis*	1	3
Corynebacterium	*species*	1	3
Enterococcus	*faecalis*	42	131
Enterococcus	*faecium*	120	676
Enterococcus	*gallinarum*	2	13
Listeria	*grayi*	2	6
Listeria	*innocua*	2	6
Listeria	*monocytogenes*	24	78
Listeria	*seeligeri*	1	3
Listeria	*welshimeri*	2	6
Micrococcus	*luteus*	5	15
Staphylococcus	*aureus*	297	1016
Staphylococcus	*capitis*	14	42
Staphylococcus	*caprae*	1	3
Staphylococcus	*carnosus*	1	3
Staphylococcus	*cohnii*	11	33
Staphylococcus	*epidermidis*	19	61
Staphylococcus	*equorum*	1	3
Staphylococcus	*haemolyticus*	11	33
Staphylococcus	*hominis*	15	48
Staphylococcus	*lugdunensis*	7	21
Staphylococcus	*saprophyticus*	8	24
Staphylococcus	*warneri*	8	24
Staphylococcus	*xylosus*	1	3
Staphylococcus	*species*	3	9
Streptococcus	*agalactiae*	7	24
Streptococcus	*anginosus*	4	18
Streptococcus	*mitis-oralis group*	5	15
Streptococcus	*pyogenes*	11	33
Streptococcus	*salivarius*	6	18
Candida	*albicans*	25	85
Candida	*glabrata*	12	51
Candida	*guilliermondii*	1	6
Candida	*krusei*	1	5
Candida	*lusitaniae*	2	6
Candida	*parapsilosis*	6	18

(continued)

Genus	Species	Isolates in Database	Spectra in Database
Candida	*tropicalis*	2	6
TOTAL		**1097**	**4026**

3.3.6 Identification and confidence level determination for identified isolates.

At each level within the spectral database as per Figure 3.2, the identification results were obtained based on the results produced for each replicate spectrum acquired from every isolate. The isolates were given an identification to one of the categories present at the tested tier within the database, or as not-identifiable based on the criteria developed (Table 3.3). Non-identifiable results were reported in cases when i) triplicate spectra did not provide the same identification result, ii) one or more spectra of the isolate predicted to be more than one category in the binary analyses, iii) spectra were not predicted to be any of the categories in the binary analyses, iv) binary analyses and multiclass analyses had contradicting identification results. Identification was only given when binary and multiclass analyses provided the same result. For isolates that obtained an identification, a newly calculated probability of the isolate belonging to the identified group was calculated based on the weighed ratio as per Table 3.4. The level of confidence was then determined based on which calculated probability range the isolates obtained (Table 3.5).

Table 3.3 Determination of isolates with correct identification, no identification or misidentification based on binary and multiclass SVM analyses

Result Outcome	Conditions
Identification results provided	Predicted result from binary and multiclass SVM are consistent
No identification	Predicted result from binary and multiclass SVM are inconsistent

Most likely identification is defined as the class with the largest probability. Based on the SVM modelling, rare cases of contradicting results between predicted and most likely identification were observed. However, predicted results were taken to be identification results, rather than most likely results in these cases.

Table 3.4 Calculation method for determining confidence levels of identified isolates

Condition	Probability of the identification result (%)	Weight given
Predicted and most likely result from SVM analysis are consistent	>50	0.5
Predicted and most likely result from SVM analysis are inconsistent	>50	0.3
Predicted and most likely result from SVM analysis are inconsistent	<50	0.3

To calculate the combined probability based on binary and multiclass SVM analyses, the probabilities determined by each analyses were taken and ratioed based on whether the predicted and most likely results were concordant or not. For cases when they were concordant, the weight of the probability calculated was given a 0.5 weight, while discordant results resulted in 0.3 weight. The relative importance of binary and multiclass SVM analyses were proportioned to be equal, with 50 % weight each.

Table 3.5 Determination of confidence level, based on the calculated probability of isolates

Calculated Probability (%) based on average from replicate spectra	Confidence Level
0-50	Low
51-80	Mid
81-100	High

Confidence level was determined based on the average of calculated probability, using the ratios mentioned in Table 3.5. Identification was determined to be low if the average probability was 50 % or lower, mid-confidence in 51-80 % range, and high for 81 % and above. Non-identified replicates were used as part of the averaging calculating with 0 % probability.

3.4 Results and Discussion

3.4.1 Comparison of transflection spectral acquisition mode against ATR and transmission modes

Three spectral acquisition modes, transmission, ATR and transflection, were compared in this study, with the focus on determining whether transflection spectral acquisition mode is

appropriate for microbial studies by FTIR spectroscopy. A summary of the notable differences between the spectral acquisition modes are listed in Table 3.6.

3.4.1.1 Comparison in sample preparation and spectral acquisition

Prior to spectral acquisition, samples required preparation for transmission and transflection modes. ATR mode required no sample preparation, and therefore was the simplest in regards to pre-spectral acquisition step. Transmission and transflection both required that isolated colonies are picked and deposited as thin films onto IR transparent and reflective materials respectively. This step may be cumbersome, depending on the microorganism and media used to culture the samples. For example, *Escherichia coli* cultured on MacConkey agar is more difficult to deposit evenly thin films compared to when cultured on blood agar due to the bile salt precipitation (caused by fermentation of lactose, present in the medium, and the decrease of pH of the agar) [27]. Optimal film thickness is required for transmission and transflection modes (2-8 µm and 1-4 µm respectively [10]). The added benefit for using transmission and transflection is that when spectral acquisition is automated with a motorized staged, manual labor required in spectral acquisition will be reduced to sample deposition and logging sample positions into a computer system, similarly to how MALDI-TOF MS systems currently function for clinical microbial identification. The ATR-FTIR spectral acquisition protocol developed by the McGill IR Group required that spectra are acquired from a wet-microbial colonies, rather than waiting for deposited samples to dry on the ATR crystals [13]. ATR spectral acquisition required sufficient biomass to cover the ATR-crystal, and rapid spectral acquisition to prevent evaporation. This may be difficult for microorganisms with slow growth (low biomass) or for those that dry into films quickly. However, it could be combatted by implementing a high moisture chamber over the sample and ATR crystal, during spectral acquisition. The disadvantage of ATR-mode is that each spectral acquisition requires deposition of new colonies, and therefore overall more biomass is required compared to transmission and transflection mode.

Spectral acquisition by ATR and transflection mode on the Cary 630 (single element DTGS detector) with 64 coadded scans and 8 cm^{-1} resolution, required approximately 1 minute per spectrum, totaling 3 minutes per isolate in this study. The FPA-FTIR spectrometer required longer set-up time compared to a single element detector spectrometer. The MCT detector on the FPA-FTIR spectrometer must be cooled down first, by filling it with liquid nitrogen. The cooled MCT

detector must maintain the cooled temperature, in order to acquire spectra with high signal to noise ratio for the FTIR images [28]. Additionally, sample acquisition by FPA-FTIR spectrometer requires the samples to be spotted and focused onto, prior to spectral acquisition through the attached microscope. Overall FTIR images require more time for spectral acquisition, compared to single element detectors. However, considering that in the time taken to obtain an FTIR image (~4 minutes), many more FTIR spectra (16,834 spectra per image in the case of the spectrometer used in this study) are generated compared to 1 ATR- and transflection-FTIR spectrum per acquisition, the FPA-FTIR excels in spectral acquisition efficiency in terms of spectrum acquired per minute. It should also be kept in mind that with each FPA-FTIR image acquisition, the data file is much larger, at approximately 114MB in this study, compared to 8-29KB per spectrum acquired by SED spectrometers. The number of datapoints generated and the file size of each spectrum is dependent on both the resolution and zero-filling. This was reflected in the transflection data files, which increased from 8 to 29KB through the increased zero-filling from 2 to 8 orders (data not shown). Thus, in order to process high volumes of FTIR images, a powerful computing system with large storage locations are required. It should also be noted that for any spectral acquisition mode, on any spectrometer, there is a correlation between acquisition time and the resolution, as well as the number of coadded scans. The higher the resolution, and the higher the number of coadded scans, the longer the spectral acquisition time [29].

Accurate microbial identification by FTIR spectroscopy relies on successful isolation of organisms prior to data acquisition. Identification of organisms rely on the spectral similarity to those part of the spectral database, and therefore, spectra acquired from a mixed culture would struggle to be identified correctly. Since spectra acquired by single element detectors acquire data from the available field of view/aperture, it is critical that the deposition is homogeneous (from one species/strain type). Perhaps the main advantage of using FPA-FTIR imaging is in its ability to identify mixed cultures. This was demonstrated by the McGill IR Group through the development of PLS models, to identify the relative proportions of different microorganisms that comprised of the FTIR images [8].

Table 3.6 Spectral acquisition parameters for FTIR spectroscopy

Parameters	**Transmission FPA-FTIR imaging**	**Transflection FPA-FTIR imaging**	**ATR**	**Transflection**
Sample preparation	Deposition onto IR transparent material (ZnSe, CaF_2)	Deposition onto IR reflective material (E-glass)	none	Deposition onto IR reflectance material (E-glass)
Spectrometer used	Cary 670 FTIR spectrometer + Cary 620 FTIR microscope	Cary 670 FTIR spectrometer + Cary 620 FTIR microscope	Cary 630 + diamond ATR accessory	Cary 630 + 10° specular reflectance accessory
Detector type	MCT (mercury cadmium telluride)	MCT	DTGS (deuterated triglycine sulfate)	DGTS
Field of View	700 x 700 µm Pixel size 5.5 x 5.5 µm	700 x 700 µm Pixel size 5.5 x 5.5 µm		6 mm aperture
Spectral Range (cm^{-1})	3900-800 (ZnSe) 3900-850 (CaF_2)	3900-800	4000-650	4000-650
Resolution (cm^{-1})	8	8	8	8
Zero-filling	Auto	Auto	2	2 + 8
Data spacing	3.8568	3.8568	1.8637	1.8637 (2 zero fill) 0.4659 (8 zero fill)
Background scan	128	128	64	64
Co-add scans	64	64	64	64
Approximate spectral acquisition time per spectrum/image	2-5 mins	2-5 mins	1 min	1 min
Post Acquisition Processing	Pixel Filtration Averaging – per image	Pixel Filtration Averaging – per image	Quality check – remove unfit spectra	Quality check – remove unfit spectra

3.4.1.2 Comparison in spectral quality check and processing

Upon spectral acquisition, the quality of each spectrum was checked prior to further analysis. The quality check process was applied to both FTIR images and spectra. For the spectral images, individual spectra that make up each acquired FTIR image were subjected to the quality check. Any spectra with high levels of noise (root-mean-square (RMS) >0.001) and spectra with absorbance outside of the 0.4-1.2 units in the amide I region (1700-1600 cm^{-1}) after baseline correction were filtered out. Figure 3.3 shows relative absorbance intensity across an FTIR image by color, with red being highest intensity, to dark blue indicating low intensity in the amide I region. The collection of spectra that passed the quality check were used to produced one average spectrum per image, which was then further used for comparison in spectral quality and discriminatory analysis. For ATR-FTIR spectra, spectra with a ratio between water (absorbance at 3200 cm^{-1}) and amide I region less than 1.33 was removed, due to the lack of moisture content in the sample during spectral acquisition. Figure 3.4 shows an example of acceptable (red) and non-acceptable (green) ATR-FTIR spectra. For transflection FTIR spectra, spectra with absorbance outside of the 0.4-1.2 absorbance units in the amide I region were removed. Figure 3.3 shows an example of acceptable (red) and non-acceptable (green) transflection FTIR spectra, based on the amide I region absorbance intensity.

Figure 3.3 Example of an FPA-FTIR image acquired from a bacterial deposition on ZnSe, showing the differences in relative absorbance intensity in spectral region 1480-980 cm^{-1} (higher absorbance is visualized as red/orange regions, and low absorbance regions are displayed as green/blue)

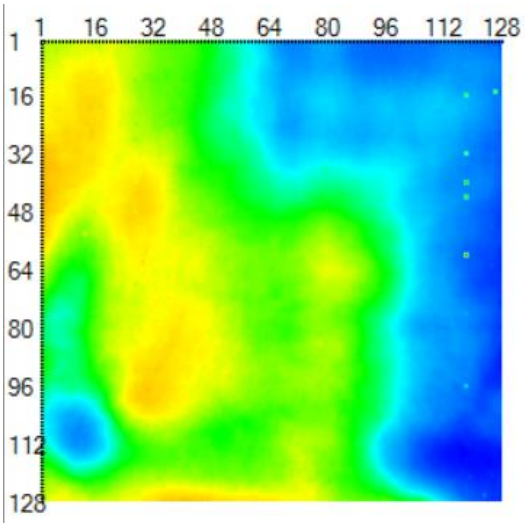
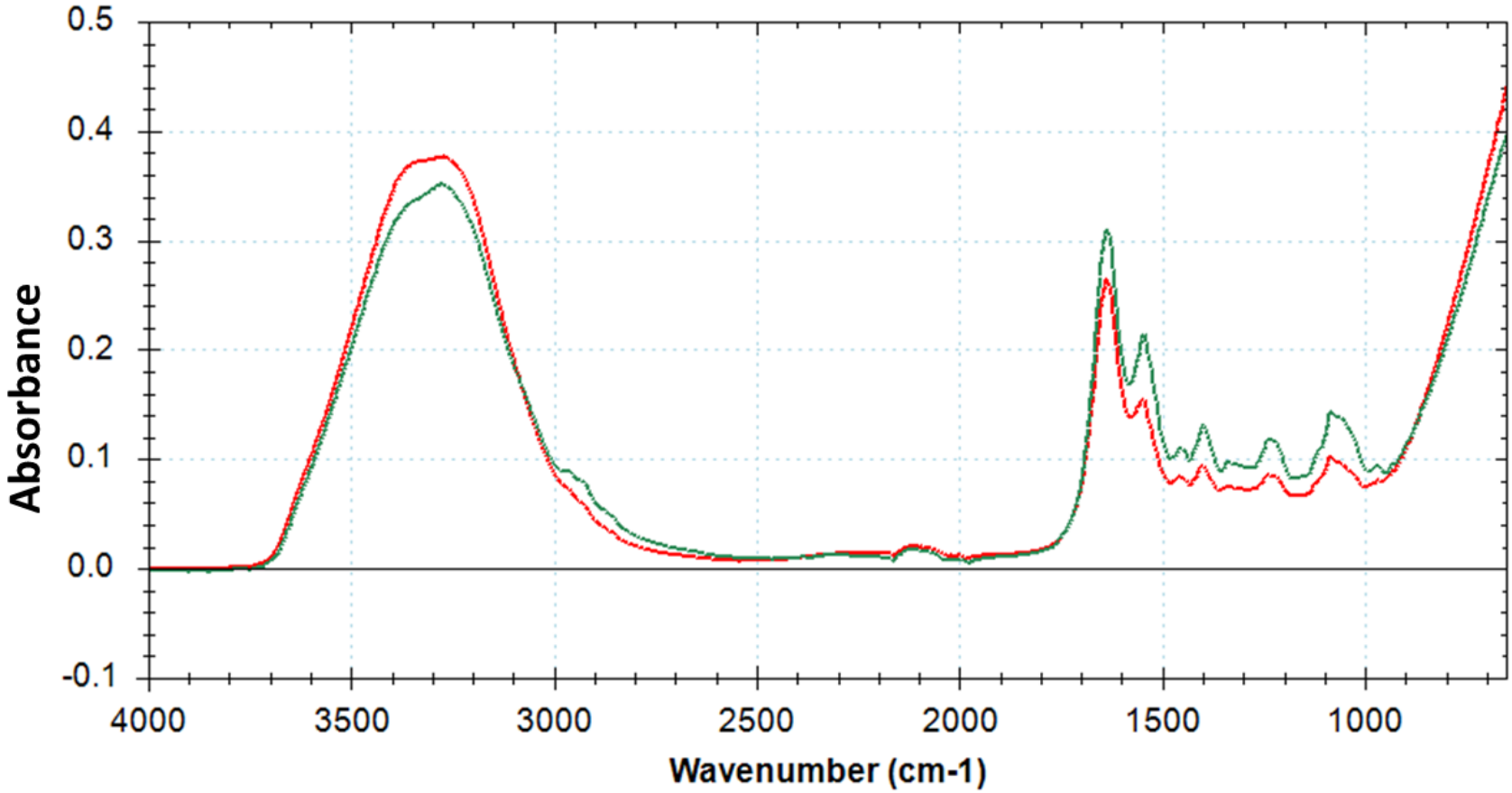
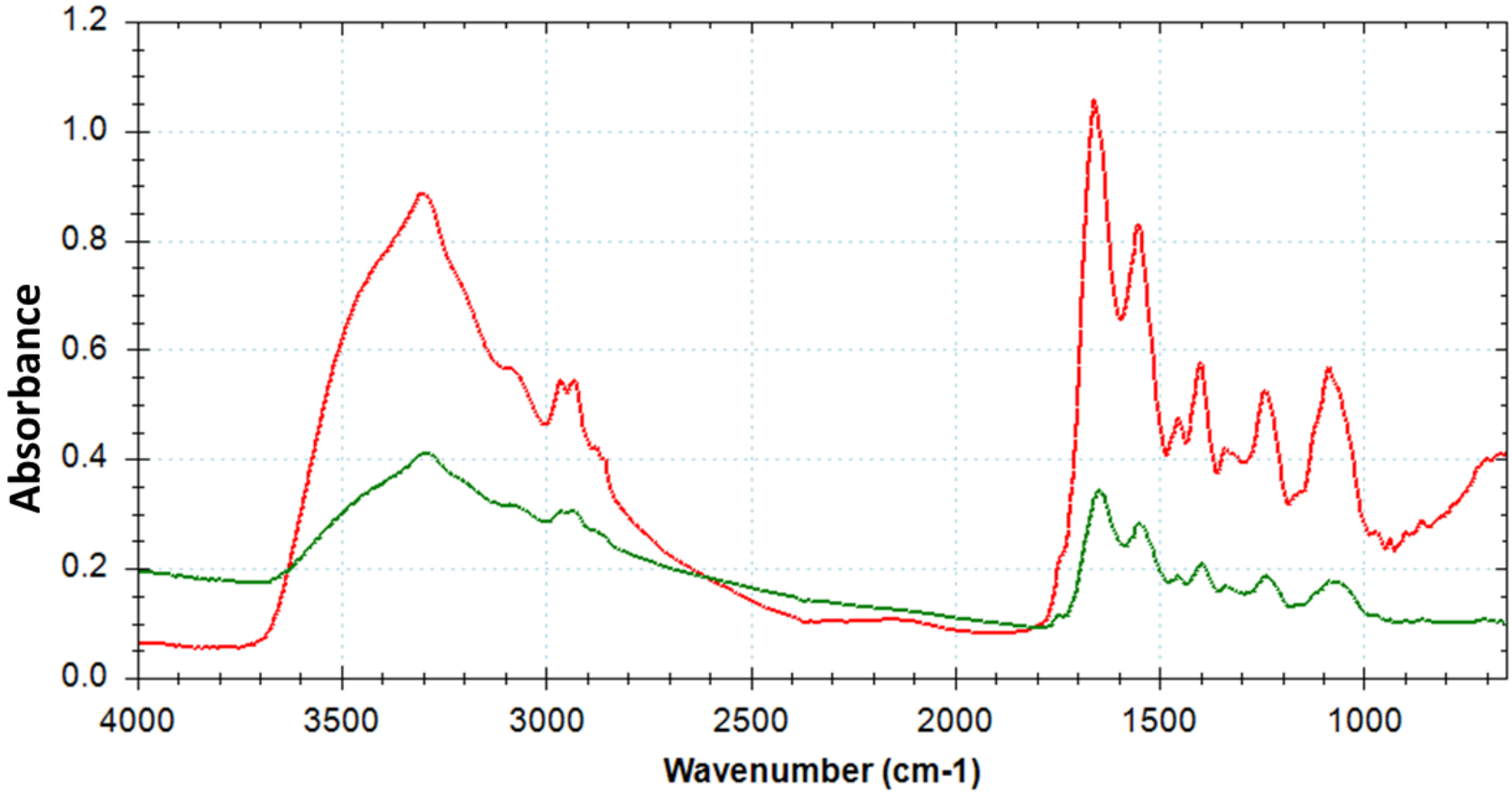


Figure 3.4 Example of good (red) and poor (green) quality ATR-FTIR spectra acquired from bacterial colonies



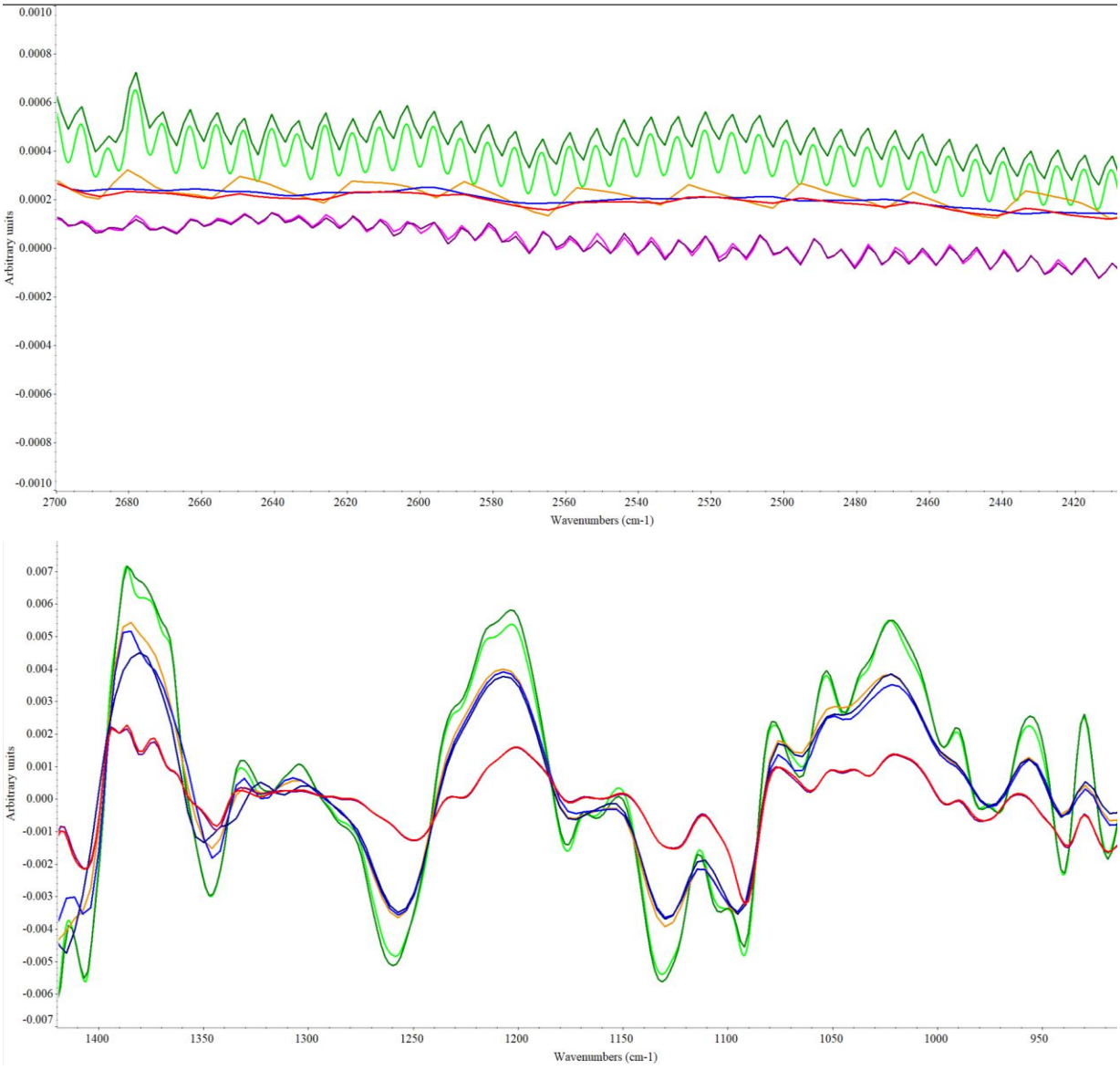
Red spectrum is an example of a passing spectrum based on the quality control parameters set. The moisture content is maintained, and the ratio between the water absorbance at 3270 cm^{-1} was 0.37, and the absorbance at amide I was 0.27, resulting in a water/amide I region ratio of 1.33. Green spectrum is an example of a poor quality spectrum due to the partial drying in the sample, represented by lower moisture-related absorbance in the 3300 cm^{-1} region, and higher absorbance observed in the amide I and II region. The ratio becomes <1.33, thereby deeming the spectra unfit for further analysis.

Figure 3.5 Example of good (red) and poor (green) quality transflection FTIR spectra acquired from bacterial depositions on E-glass



Red spectrum is an example of a passing spectrum based on the quality control parameters set. The absorbance at amide I was 1.06. Green spectrum is an example of a poor quality spectrum due to the low overall absorbance, relating to insufficient sample deposition. The lack of sample on the E-glass resulted in baseline shift. Such spectra were removed from further analysis.

Figure 3.6 Overlay averaged, first derivative FTIR spectra acquired by transmission FPA (ZnSe in orange, CaF_2 in blue), ATR (pink and purple), and transflection (light and dark green for single element DGTS and red for FPA) in region 2700-2400 cm^{-1} and 1400-900 cm^{-1}.



Noise level of averaged, first derivative FTIR spectra acquired from transmission (by ZnSe in Orange, CaF_2 in blue), transflection (light and dark green on single element DGTS detector with 8 and 2 zero fill respectively, FPA in red), and ATR (pink and purple) in spectral region 2700-2400 cm^{-1}. Spectral region 1400-900 cm^{-1} shows the signal in first derivative spectra.

Table 3.7 Calculated noise level comparison in spectra acquired by different acquisition modes

Spectral Acquisition Mode	**Number of Averaged spectra**	**Noise (Peak to Peak) 2700-2400 cm^{-1}**	**Noise (root mean squared) 2700-2400 cm^{-1}**	**Signal to Noise Ratio (SNR) Mean/SD (whole spectrum)**
ATR	72	0.0001154	2.991×10^{-5}	0.874991
ATR	36	0.0001248	3.004×10^{-5}	0.869314
ATR	36	0.0001279	3.22×10^{-5}	0.882867
Transflection	36	0.0003683	7.153×10^{-5}	1.450002
Transflection 2 zerofill	36	0.0003367	6.21×10^{-5}	1.579071
E-glass FPA	48*	7.056×10^{-5}	1.694×10^{-5}	1.124291
ZnSe FPA	35*	0.0001698	3.706×10^{-5}	0.857766
CaF_2 FPA	28*	5.65×10^{-5}	1.437×10^{-5}	1.364181

Noise levels were calculated from averaged spectra (number of spectra used to generate the average spectrum per acquisition mode are indicated in the second column of Table 3.7), upon application of vector normalization and first derivative. In Figure 3.6, ATR spectra (pink and purple) displayed smaller noise level, compared to transflection (dark and light green). FPA-FTIR spectrometer acquired spectra (transmission on ZnSe and CaF_2 and transflection on E-glass) had even smaller noise levels. However, taking into consideration that these averages are obtained from averaged spectra taken from each FTIR image acquired (hundreds to tens of thousands of spectra averaged into one), the noise per acquired spectrum by FPA-FTIR spectrometers, regardless of the acquisition mode, is inferior to those acquired by single-DTGS detectors. Due to the inherent non-uniformity between detector elements in the FPA, the noise level is higher in FPA detectors compared to single element detectors. Further adding to the problem, the repeated temperature cycling of array detectors can result in pixel degradations and inoperative pixels over time [28, 30]. Signal-to-noise ratio (SNR) was calculated by the mean signal detected across the whole spectrum over the standard deviation. Although ATR has low noise level, the SNR was lower than transflection because ATR has a smaller pathlength compared to transflection, which results in overall lower signal (observed in Figure 3.6) [10].

.

3.4.1.3 Comparison in spectral discrimination capabilities

For each spectral acquisition mode, analysis in discriminating FTIR spectra of Gram-negative bacteria from Gram-positive bacteria was conducted, and shown to be successful in a dendrogram generated by HCA (Figure 3.8), and PCA (transmission spectral data example shown in Figure 3.7. Additionally, successful spectral discrimination between *S. aureus* and coagulase negative staphylococcal species (CoNS) was demonstrated for each spectral acquisition mode upon identifying optimal spectral regions that enhance the differentiation between the two groups, using forward region selection algorithms (Figure 3.8). Due to the limited sample size, no further analyses could be conducted to evaluate whether transflection mode could provide FTIR spectra for microbial analysis, in direct comparison to transmission and ATR modes. At this level of discrimination, transflection spectral acquisition mode showed discriminatory capabilities that matched the ATR and transmission FTIR data that were previously published for *S. aureus* discrimination from CoNS [26, 31]. To provide further in-depth analysis and comparison between the acquisition modes on their discriminatory capabilities, the three acquisition modes should be compared directly and evaluated on their abilities to discriminate at the sub-species and strain typing levels.

Figure 3.7 Spectral discrimination of averaged transmission spectra acquired on ZnSe and CaF_2 by genera of select Gram-negative (purple) and Gram-positive bacteria (blue, green and red) showcased successful spectral discrimination in a PC score plot for PC3 vs PC2

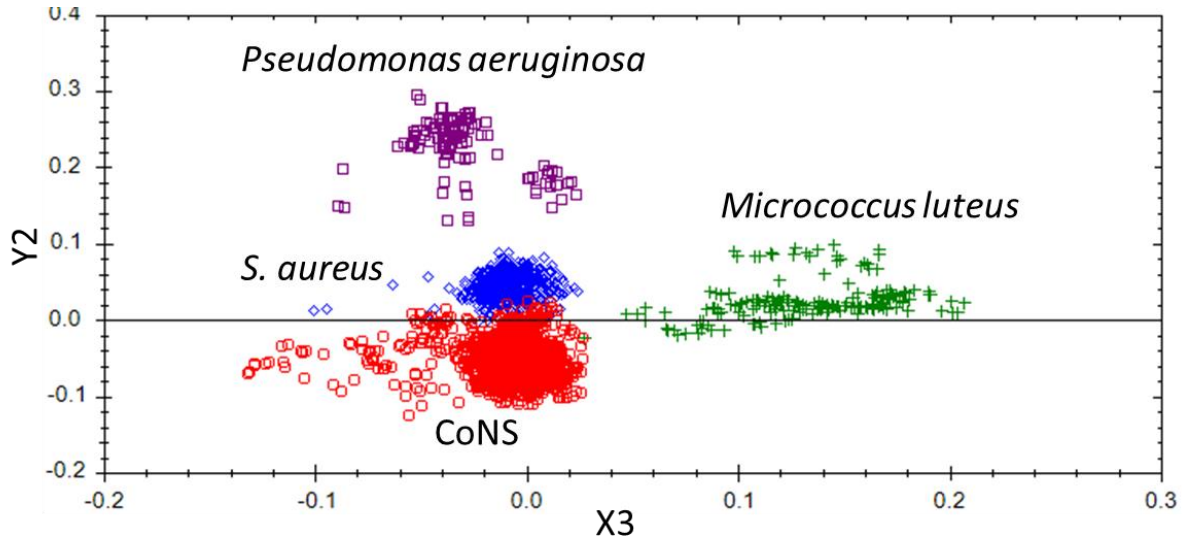
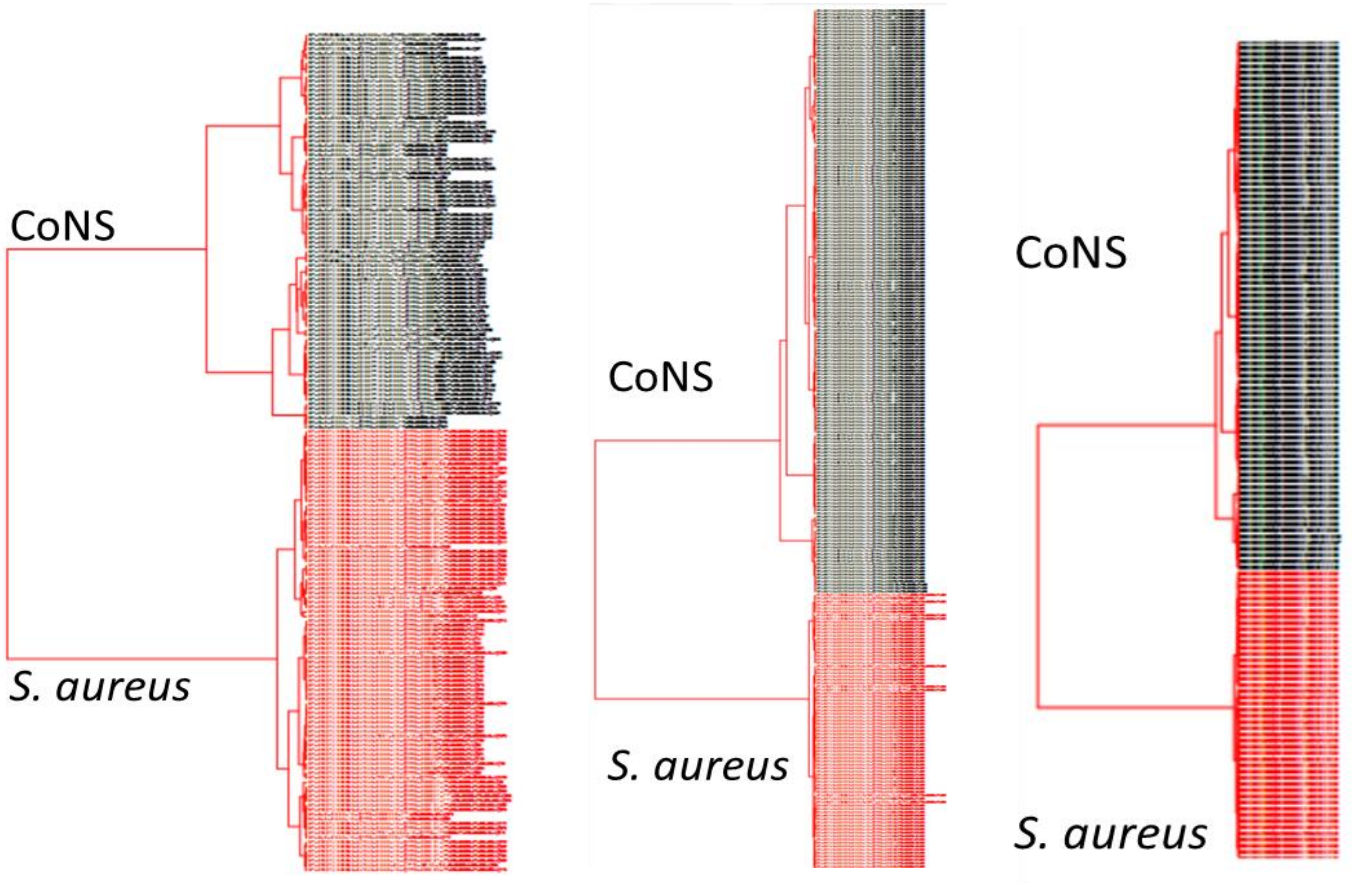


Figure 3.8 Dendrogram generated by HCA visualizing the discrimination between S. aureus (Red) and CoNS (Black) using spectral regions selected from feature selection algorithm on FTIR spectra acquired in a) transmission mode on ZnSe from averaged FTIR images, b) ATR- and c) transflection acquisition modes.



3.4.2 Transflection FTIR spectral database construction and evaluation.

To further evaluate the discriminatory capabilities of transflection FTIR spectral data for microbial analysis, a multitier spectral database was constructed, and tested for identification of microbial isolates. Using PCA-SVM, the isolates used for spectral database construction were successfully discriminated at each level, with correct prediction ranging from 86.1 - 100 % for the training set (Table 3.8 - Table 3.13). The database was created by taking 60 % of the isolates from retrospective sets of data acquired at clinical and food microbiology laboratories, with a particular focus on *Enterococcus* and *Staphylococcus* species. *Staphylococcus* and *Enterococcus* species isolates represented 36.2 % (*n* = 396) and 15.1 % (*n* = 165) of the spectral database respectively. On the Gram-negative bacteria side, *Salmonella* (*n* = 97, 8.9 %) and *Escherichia* species (*n* = 59, 5.4 %) had the largest representation. Genera and species that were chosen to be represented required to have 10 or more isolates must be included in the database, with the exception of *Corynebacterium* (*n* = 9), *Micrococcus* (*n* = 5), and *Acinetobacter* (*n* = 8) which had unique FTIR spectral characteristics that enabled for correct identification with little spectral representation. *Serratia* species (*n* = 8), were included in the database due to its clinical importance [32, 33]. At the species level, databases were constructed for select species within the genera *Enterococcus* (*E. faecalis, E. faecium, E. gallinarum*), *Staphylococcus* (*S. aureus, S. capitis, S. cohnii, S. epidermidis, S. haemolyticus, S. hominis, S. lugdunensis, S. saprophyticus, S. warneri*) and *Shigella* (*S. flexneri* and *S. sonnei*). Additionally, due to the importance of discriminating and identifying *Listeria monocytogenes* from non-monocytogenes *Listeria* species, the database was created to discriminate between the two groups [34]. Misidentification and no-identification rates ranged from 0-4.4 % and 0-9.5 % respectively for the training set across all levels of the multi-tier spectral database. A total of 37 isolates had no-identification, and 17 isolates were mis-identified at the genus level. Upon inspection of the incorrect and non-identified isolates in the training set, at the genus level, all errors occurred within the Gram-negative bacterial isolates (Table 3.11). Of the 14 Gram-negative genera that were represented in the database, 6 genera (*Achromobacter, Acinetobacter, Klebsiella, Proteus, Serratia, Shigella*) had > 90 % correct discrimination in the training set. Five genera (*Burkholderia, Escherichia, Pseudomonas, Salmonella,* and *Stenotrophomonas*) had a correct discrimination rate ranging between 80 % and 90 %. Gram-negative genera that had less than 80% correct discrimination in the training set were *Aeromonas, Citrobacter,* and *Enterobacter* with correct discrimination rates of 64. 3%, 52.4 % and 68.2 %

respectively, contributed to 7 misidentified and 15 non-identifiable isolates combined. *Escherichia* and *Salmonella* also contributed 6 and 11 isolates that could not be identified respectively. At the species level, the training model resulted in 98.2 % (*n* = 606) correct classification (Table 3.12). The 6 non-identified isolates were *S. cohnii* (*n* = 1), *S. lugdunensis* (*n* = 3), *S. saprophyticus* (*n* = 1), and *S. warneri* (*n* = 1). With the exception of *S. lugdunensis,* and *S. warneri*, > 90 % correct discrimination at the species level was achieved. 100 % discrimination between *E. faecalis* and *E. faecium, S. flexneri* and *S. sonnei*, and *L. monocytogenes* and non-monocytogenes *Listeria* species were achieved. The constructed database showed that discrimination of microbial species were achievable using transflection-FTIR spectra when employing PCA-SVM.

Using the developed database, the 40 % of isolates that were set aside as validation set were identified. 99.9 % correct identification was achieved at the Gram-level, and all yeast isolates were correctly identified (*n* = 34). At the genus level, the correct identification rates were 72.8 % (*n* = 182) and 98.9 % (*n* = 447) respectively for Gram-negative and Gram-positive isolates. 0.9 % (*n* = 4) and 0.2 % (*n* = 1) of Gram-positive isolates were misidentified and not-identified respectively, whereas 9.2 % (*n* = 23) of Gram-negative bacterial isolates were misidentified and 18.0 % (*n* = 45) were not identified. Errors in the Gram-positive isolates were due to one *Corynebacterium* isolate not being identifiable past the Gram-level, and misidentification of 2 *Bacillus*, 1 *Enterococcus* and 1 *Listeria* isolates. The non-identified *Corynebacterium* isolate had low probability (10-25 % probability) of it being *Bacillus, Corynebacterium* or *Micrococcus* in the binary analyses, and less than 50 % probability it being *Corynebacterium* in the multiclass SVM analysis. PCA revealed that the FTIR spectra of this isolate was spectrally distinct from the other *Corynebacterium* isolates (Figure 3.9). One *Bacillus* isolate was predicted to be *Staphylococcus* species and another predicted as a Gram-negative bacteria. These isolates were also individually re-analysed, and based on the characteristics in the spectral range 1350-800 cm^{-1}, spectra of these two *Bacillus* isolates were more similar to *Staphylococcus* and Gram-negative bacteria (Figure 3.10 - Figure 3.13).

Achromobacter, Acinetobacter, and *Shigella* achieved > 90 % correct identification in the validation set. However, most other genera had poor correct identification rates below 80 %. In particular, *Citrobacter* had 18.2 % (*n* = 2) and *Enterobacter* had 40.0 % (*n* = 6) correct

identification. Genera within the Enterobacteriaceae family such as *Citrobacter*, *Enterobacter*, *Escherichia*, *Klebsiella* and *Salmonella* had many non-identifiable isolates, making up 34 of 37 non-identifiable isolates. Out of these genera, *Escherichia*, *Klebsiella* and *Salmonella* had the largest number of isolates present in the database, with 59, 41 and 97 respectively. These were the genera with the highest correct identification rates (68.9 %, 84.6 % and 82.0 % respectively), with the exception of *Shigella* (100 % correct) which was represented by 25 isolates.

At the species level, 95.2 % of the isolates were correctly identified in the validation set. Fourteen isolates (3.3 %) were not identified, and 6 isolates (1.4 %) were misidentified. 100% correct identification was achieved for *E. faecalis, E. faecium, S. capitis, S. cohnii, S. flexneri* and *S. sonnei*. Additionally, *L. monocytogenes, S. aureus, S. epidermidis* had > 90 % correct identification (Table 3.12). However, some species identification performed poorly. These included *S. haemolyticus, S. hominis, S. saprophyticus* and *S. warneri* with 44.4 % (n = 4), 75.0 % (n = 9), 57.1 % (n = 4) and 50.0 % (n = 4) correct identification respectively. The common factor among these species was that they each had 15 or less isolates representing the species in the spectral database, indicating a lack of representation in the spectral database. The discordant results and non-identified isolates within the CoNS were also observed at the cluster group level (Table 3.13). This was especially the case for the haemolyticus group, saprophyticus group and warneri group. Lugdunensis, saprophyticus and warneri groups especially have low representation in the database, with 7, 11, 8 isolates respectively.

The test set constituted of isolates collected over a four year period (2017-2020), in four different microbiology laboratories, located in Montreal, Canada, Winnipeg, Canada and Brisbane, Australia (Table 3.2). A set of isolates were obtained at CHUSJ in the following months after isolates were collected for database construction and validation. Isolates collected at MUHC mainly focused on isolates that were later used for strain typing, for nosocomial outbreak detection. Isolates collected in Brisbane, Australia were derived from skin swabs, obtained from patients with different stages of skin cell carcinoma. Isolates collected at NML were cultured on modified blood agar plates. Overall, there were 1793 isolates that were used to evaluate the database performance. Fifteen isolates were yeast and 1778 isolates were bacteria. 80 % (n = 12) of yeast isolates were correctly identified as yeast, with 3 isolates incorrectly predicting as *Staphylococcus* species. All 1606 bacterial isolates predicted correctly as bacteria. At the Gram-level, 95.9 % (n = 260) and 98.8 % (n = 1491) concordant identification was achieved for Gram-negative and Gram-positive

bacteria respectively. Within the Gram-positive bacteria, 95.8 % (*n* = 1443) correct genus identification was achieved, with 0.9 % (*n* = 14) and 3.3 % (*n* = 50) of the isolates resulting in no-identification and misidentification respectively. Of the 50 misidentified isolates, 16 were misidentified at the Gram-level, where the Gram-positive isolates were identified as Gram-negative bacteria. The majority of the misidentification at the genus level within Gram-positive bacteria resulted from non-represented genera isolates being identified as one of the genera represented in the database. For example, *Lactobacillus* (*n* = 3, 75 %) and *Lactococcus* isolates (*n* = 1, 100 %) predicted as *Enterococcus*, a close genus within the lactic acid bacteria group [35]. Overall, of the 21 non-represented isolates that were part of the test set, 4 isolates (19.0 %) could not be identified past Gram-positive, and the remaining 17 isolates (81.0 %) were misidentified as one of seven represented genera in the database.

Amongst the 7 Gram-positive genera that were represented in the database, *Bacillus* performed the worst, with 27.3 % (*n* = 3) correct identification, and 72.7 % (*n* = 8) misidentification. *Bacillus* isolates were predicted to be GN (*n* = 4), *S. aureus* (*n* = 3) and *Corynebacterium* (*n* = 1). All other genera that were represented in the database (*Corynebacterium*, *Micrococcus*, *Enterococcus*, *Staphylococcus* and *Streptococcus*) had high concordant identification results, ranging between 87.2 and 100 %. There were no *Listeria* isolates within the test set to evaluate the database for its identification performance. At the species level, 90.6 % (*n* = 1185) correct identification was achieved, with 3.7 % (*n* = 49) no identification and 5.7% (*n* = 74) mis-identification rates. Identification of *E. faecalis* and *E. faecium* was highly concordant to reference methods, at 98.5 % (*n* = 65) and 99.2 % (*n* = 360) respectively. Other enterococcal species (*E. avium, E. casseliflavis* and *E. gallinarum*) had low representation or no representation in the database, and therefore these isolates were mostly incorrectly identified as *E. faecalis*, or not identified at the species level. *S. aureus* had a 98.4 % (*n* = 663) correct identification, showcasing high concordance between transflection-FTIR spectroscopy and reference methods. However staphylococcal species identification within the coagulase negative staphylococcus species did not perform as well, ranging between 0 % and 100 % correct species identification. *S. cohnii, S. lugdunensis* and *S. warneri* had no correct identification. *S. saprophyticus* had 100 % (*n* = 1) correct identification. Combined with the 7 isolates tested in the validation set, 5 out of 8 *S. saprophyticus* isolates (62.5 %) predicted correctly using this database. Although species like *S. capitis, S. cohnii, S. epidermidis*, and *S. hominis* had > 90 % correct identification in the

validation set, the performance using the test set was much worse, with 58.3 %, 0 %, 61.6 %, and 71.0 % concordant results respectively, relative to reference methods. Species such as *S. capitis*, *S. lugdunensis* and *S. saprophyticus* had contrasting performance results between validation and test sets. *S. saprophyticus* had 57.1 % (*n* = 4) correct identification, and *S. lugudnensis* had 100 % (*n* = 1) correct identification in the validation set. But both species had 0% correct identification in the test set. *S. capitis* performed well in the validation set but not in test set, with 100 % (*n* = 8) and 62.5 % (*n* = 15) correct identification respectively. Despite having 14 isolates representing the species in the database, accurate species identification could not be achieved. Of the 6 misidentified *S. capitis* isolates, 3 isolates were identified as *S. epidermidis,* another species within the *S. epidermidis* subgroup [36]. Another 3 isolates were identified as *S. haemoylticus,* and *S. hominis,* species within the *S. haemolyticus* subgroup [36]. Other species such as *S. caprae, S. pasteuri,* and *S. xylosus* were also tested, despite their lack of representation in the database. 62.5 % (*n* = 10) of non-represented *Staphylococcus* species achieved no-identification at the species level, with the remaining 37.5 % (*n* = 6) being misidentified as one of the species present in the database.

The successful identification of *S. aureus* (*n* = 216, 99.5 % and *n* = 663, 98.4 % respectively for validation and test set) indicated that the database is robust and is able to identify *S. aureus* from CoNS species with high confidence, due to sufficient spectral representation in the database for *S. aureus*. Since *S. aureus* is clinically more important compared to other *Staphylococcus* species, it is a crucial component for the spectral database to achieve [37]. In addition, the successful identification of *E. faecalis* and *E. faecium* isolates showcase that with sufficient spectral representation in the database, isolates of that genus and species can be correctly identified.

Contrary to Gram-positive bacteria identification results, Gram-negative genus identification performed poorly in the test set, with 41.3 % overall correct identification (*n* = 112), 28.0 % (*n* = 76) no identification and 30.6 % (*n* = 83) misidentification rates. While some genera like *Proteus* (*n* = 4) and *Achromobacter* (*n* = 6) achieved 100 % correct identification, other genera that were represented in the database were not able to identify isolates correctly. In particular, *Aeromonas* (*n* = 6), *Burkholderia* (*n* = 3), *Citrobacter* (*n* = 1) had no correct identification, while *Acinetobacter* (*n* = 9), *Enterobacter* (*n* = 23), *Escherichia* (*n* = 46), *Klebsiella* (*n* = 17), *Serratia* (*n* = 3), *Shigella* (*n* = 10), and *Stenotrophomonas* (*n* = 18) all had less than 80 % of test isolates were correctly identified. With the exception of *Escherichia, Klebsiella* and *Salmonella,* other

genera were represented in the database with less than 25 isolates. Genera such as *Aeromonas*, *Citrobacter* and *Enterobacter* had < 80 % correct discrimination in the training set, which likely affected and contributed to the low identification performance of the test isolates. Another factor that may have contributed to misidentification is that different species within a genus may have vastly different FTIR spectral features as a result of varying biochemical and metabolic functions [38]. Therefore without sufficient species representation within each of these genera, certain species may not predict correctly when underrepresented in the database. This may have been the case for *Pseudomonas* isolates that were incorrectly identified in the test set. Of the 43 tested *Pseudomonas* isolates, 9 isolates were non-*P. aeruginosa* species, none of which were represented in the database. Seven out of these 9 isolates (77.8 %) did not predict correctly as *Pseudomonas* isolates, whereas only 2 of 34 *P. aeruginosa* isolates (5.9 %) was misidentified, and 3 (8.8 %) could not be identified at the genus level.

34 of 70 (48.6 %) non-represented Gram-negative genera in the test set were not predicted to be identifiable at the genus level, while the remaining 51.4 % (n = 36) were incorrectly identified as one of the genera present in the spectral database (Table 3.14). Compared to isolates of non-represented Gram-positive genera, more isolates were properly classified as not-identifiable at the genus level for the Gram-negative bacterial isolates. Overall 40.0 % (n = 38) of non-represented Gram-negative bacterial isolates were reported as not-identifiable. At the species level, for *Shigella* species, only two species were represented in the database. The SVM forces all interrogated isolates to be identified as one of the two groups, and does not provide a third option - "neither of the two groups". Hence, isolates were forced to be identified to the group that they are spectrally most similar to. This was the case for species such as *S. dysenteriae*, which predicted as *S. sonnei* with mid-level confidence (77.1 % probability being *S. sonnei*). Unlike *Shigella* species, *Enterococcus* species level identification was based on discrimination between 3 species, *E. faecalis*, *E. faecium* and *E. gallinarum*. The issue in the *Enterococcus* species identification was that only two isolates represented *E. gallinarum* in the spectral database, and hence, none of the tested *E. gallinarum* isolates could correctly be identified. However, the presence of a third species option opened up possibilities for isolates to be not-identifiable at the species level, rather than being forced into predicting as one of two available options.

Table 3.8 Identification results for training, validation and test set at the bacteria-yeast level

	Training							Validation							Test						
	Total	Correct (n, %)		No ID (n, %)		Mis ID		Total	Correct (n, %)		No ID (n, %)		Mis ID		Total	Correct (n, %)		No ID (n, %)		Mis ID	
Bacteria	1047	1047	100	0	0	0	0	706	706	100	0	0	0	0	1774	1774	100	0	0	0	0
Yeast	49	49	100	0	0	0	0	34	34	100	0	0	0	0	15	12	80	0	0	3	20
TOTAL	**1096**	**1096**	**100**	**0**	**0**	**0**	**0**	**740**	**740**	**100**	**0**	**0**	**0**	**0**	**1789**	**1786**	**99.8**	**0**	**0**	**3**	**0.2**

Table 3.9 Identification results for training, validation and test set for Gram-negative and Gram-positive bacteria.

Gram	**Training**							**Validation**							**Test**						
	Total	Correct (n, %)		No ID (n, %)		Mis ID (n, %)		Total	Correct (n, %)		No ID (n, %)		Mis ID (n, %)		Total	Correct (n, %)		No ID (n, %)		Mis ID (n, %)	
GN	393	393	100	0	0	0	0	254	254	100	0	0	0	0	269	258	95.9		0	11	4.1
GP	654	654	100	0	0	0	0	452	451	99.8	0	0	1	0.2	1505	1489	98.9	0	0	16	1.1
TOTAL	**1047**	**1047**	**100**	**0**	**0**	**0**	**0**	**706**	**705**	**99.9**	**0**	**0**	**1**	**0.1**	**1774**	**1747**	**98.5**	**0**	**0**	**27**	**1.5**

GN = Gram-negative. GP = Gram-positive.

Table 3.10 Gram-positive genus identification results for training, validation and test set

GP Genera	**Training**							**Validation**							**Test**						
	Total	Correct (n, %)		No ID (n, %)		Mis ID (n, %)		Total	Correct (n, %)		No ID (n, %)		Mis ID (n, %)		Total	Correct (n, %)		No ID (n, %)		Mis ID (n, %)	
Aerococcus															2	0	0	0	0	2	100
Bacillus	15	15	100	0	0	0	0	9	7	77.8	0	0	2	22.2	11	3	27.3	0	0	8	72.7
Corynebacterium	9	9	100	0	0	0	0	3	2	66.7	1	33.3	0	0	6	6	100	0	0	0	0
Dermabacter															3	0	0	0	0	3	100
Enterococcus	164	164	100	0	0	0	0	94	93	98.9	0	0	1	1.1	466	457	98.1	6	1.3	3	0.6
Gordonia															1	0	0	0	0	1	100
Kocuria															3	0	0	0	0	3	100
Kytococcus															1	0	0	1	100	0	0
Lactobacillus															4	0	0	1	25	3	75
Lactococcus															1	0	0	0	0	1	100
Listeria	31	31	100	0	0	0	0	26	25	96.2	0	0	1	3.8							
Micrococcus	5	5	100	0	0	0	0	3	3	100	0	0	0	0	11	10	90.9	1	9.1	0	0
Rhodococcus															1	0	0	1	100	0	0
Rothia															1	0	0	0	0	1	100
Staphylococcus	397	397	100	0	0	0	0	289	289	100	0	0	0	0	898	888	98.9	3	0.3	8	0.9
Streptococcus	33	33	100	0	0	0	0	28	28	100	0	0	0	0	94	79	84	0	0	15	16
Turicella															2	0	0	0	0	2	100
TOTAL	**654**	**654**	**100**	**0**	**0**	**0**	**0**	**452**	**447**	**98.9**	**1**	**0.2**	**4**	**0.9**	**1505**	**1443**	**95.9**	**13**	**0.9**	**50**	**3.3**
REPRESENTED GENERA	**654**	**654**	**100**	**0**	**0**	**0**	**0**	**452**	**447**	**98.9**	**1**	**0.2**	**4**	**0.9**	**1486**	**1443**	**97.1**	**10**	**0.7**	**34**	**2.3**

Table 3.11 Gram-negative genus identification results for training, validation and test set.

GN Genera	**Training**							**Validation**							**Test**						
	Total	Correct (n, %)		NO ID (n, %)		Mis ID (n, %)		Total	Correct (n, %)		NO ID (n, %)		Mis ID (n, %)		Total	Correct (n, %)		NO ID (n, %)		Mis id (n, %)	
Achromobacter	22	22	100	0	0	0	0	13	12	92.3	0	0	1	7.7	6	6	100	0	0	0	0
Acinetobacter	8	8	100	0	0	0	0	5	5	100	0	0	0	0	9	3	33.3	0	0	6	66.7
Aeromonas	14	9	64.3	0	0	5	35.7	7	4	57.1	0	0	3	42.9	6	0	0	0	0	6	100
Burkholderia	13	11	84.6	2	15.4	0	0	8	6	75	2	25	0	0	3	0	0	2	66.7	1	33.3
Citrobacter	21	12	57.1	9	42.9	0	0	11	1	9.1	9	81.8	1	9.1	1	0	0	0	0	1	100
Enterobacter	22	16	72.7	5	22.7	1	4.5	15	7	46.7	5	33.3	3	20	23	3	13	14	60.9	6	26.1
Escherichia	59	53	89.8	6	10.2	0	0	45	31	68.9	10	22.2	4	8.9	46	30	65.2	8	17.4	8	17.4
Klebsiella	41	39	95.1	2	4.9	0	0	26	22	84.6	3	11.5	1	3.8	17	12	70.6	2	11.8	3	17.6
Proteus	11	10	90.9	0	0	1	9.1	5	4	80	1	20	0	0	4	4	100	0	0	0	0
Pseudomonas	33	28	84.8	1	3	4	12.1	20	13	65	3	15	4	20	43	32	74.4	2	4.7	9	20.9
Salmonella	97	88	90.7	9	9.3	0	0	61	50	82	10	16.4	1	1.6	12	9	75	1	8.3	2	16.7
Serratia	8	8	100	0	0	0	0	5	3	60	1	20	1	20	3	2	66.7	1	33.3	0	0
Shigella	25	24	96	0	0	1	4	16	16	100	0	0	0	0	10	2	20	6	60	2	20
Stenotrophomonas	15	12	80	2	13.3	1	6.7	13	7	53.8	2	15.4	4	30.8	18	8	44.4	5	27.8	5	27.8
Vibrio	4	0	0	1	25	3	75	4	0	0	0	0	4	100							
TOTAL	**393**	**340**	**0.87**	**37**	**0.09**	**16**	**0.04**	**254**	**181**	**0.71**	**46**	**0.18**	**27**	**0.11**	**201**	**111**	**0.55**	**41**	**0.2**	**49**	**0.24**

Table 3.12 Species identification for Enterococcus, Listeria, Staphylococcus and Shigella isolates in training, validation and test sets

species	Training							Validation							Test						
	Total	Correct (n, %)		No ID (n, %)		Mis ID (n, %)		Total	Correct (n, %)		No ID (n, %)		Mis ID (n, %)		Total	Correct (n, %)		No ID (n, %)		Mis ID (n, %)	
E. faecalis	42	42	100	0	0	0	0	17	17	100	0	0	0	0	66	65	98.5	1	1.5	0	0
E. faecium	120	120	100	0	0	0	0	77	75	97.4	0	0	2	2.6	363	360	99.2	1	0.3	2	0.6
E. gallinarum	2	2	100	0	0	0	0	0	0	0	0	0	0	0	16	0	0	4	25	12	75
E. species	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	4	100
L. monocytogenes	24	24	100	0	0	0	0	20	17	85	2	10	1	5	0	0	0	0	0	0	0
L. species	7	7	100	0	100	0	0	6	4	66.7	0	0	2	33.3	0	0	0	0	0	0	0
S. aureus	297	297	100		0	0	0	217	216	99.5	0	0	1	0.5	674	663	98.4	0	0	11	1.6
S. capitis	14	14	100	0	0	0	0	8	7	87.5	0	0	1	12.5	24	14	58.3	3	12.5	7	29.2
S. cohnii	11	11	100	0	0	0	0	7	6	85.7	1	14.3	0	0	1	0	0	1	100	0	0
S. epidermidis	19	19	100	0	0	0	0	16	13	81.3	2	12.5	1	6.3	73	45	61.6	16	21.9	12	16.4
S. haemolyticus	11	10	90.9	0	0	1	9.1	9	5	55.6	3	33.3	1	11.1	14	4	28.6	6	42.9	4	28.6
S. hominis	15	14	93.3	0	0	1	6.7	12	9	75	2	16.7	1	8.3	31	22	71	9	29	0	0
S. lugdunensis	7	4	57.1	3	42.9	0	0	1	1	100	0	0	0	0	9	0	0	6	66.7	3	33.3
S. saprophyticus	8	7	87.5	1	12.5	0	0	7	4	57.1	0	0	3	42.9	2	0	0	2	100	0	0
S. warneri	8	7	87.5	0	0	1	12.5	8	3	37.5	3	37.5	2	25	9	0	0	7	77.8	2	22.2
S. species	4	0	0	0	0	4	100	4	0	0	0	0	4	100	18	0	0	12	66.7	6	33.3
S. flexneri	14	13	92.9	0	0	1	7.1	9	9	100	0	0	0	0	1		0		0	1	100
S. sonnei	11	11	100	0	0	0	0	7	7	100	0	0	0	0	4	1	25		0	3	75
S. species															5					5	100
Total	**614**	**602**	**98**	**4**	**0.7**	**8**	**1.3**	**425**	**393**	**92.5**	**13**	**3.1**	**19**	**4.5**	**1314**	**1174**	**89.3**	**68**	**5.2**	**72**	**5.5**

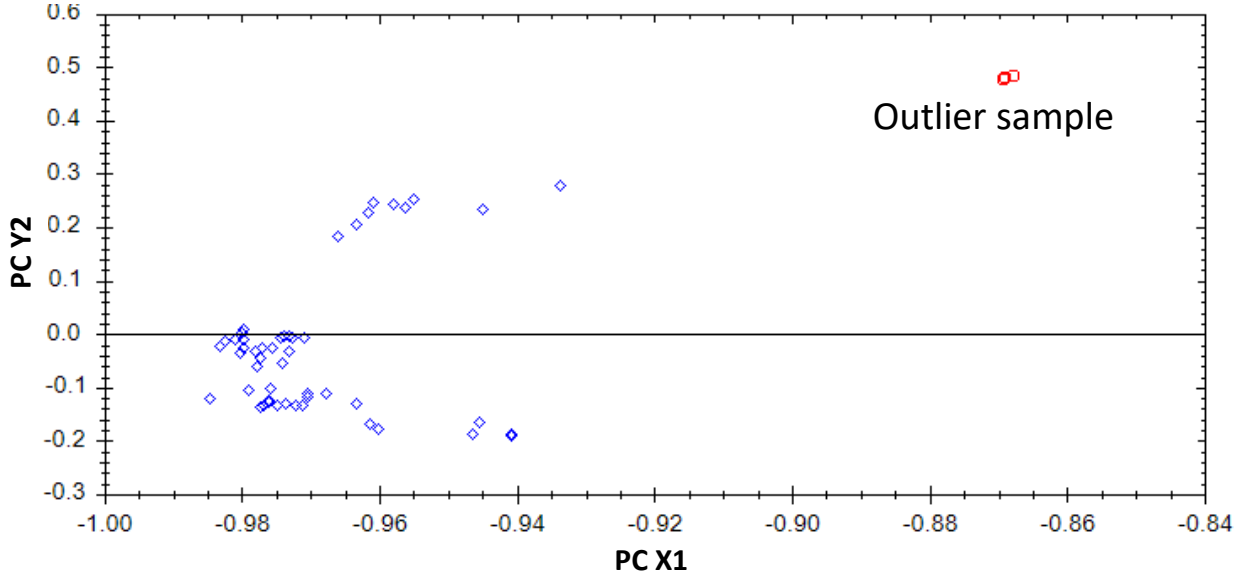
Table 3.13 Coagulase negative Staphylococcus species subgroup identification, based on phylogenetic and biochemical properties of Staphylococcus species

CoNS subgroups	**Training**							**Validation**							**Test**						
	Total	Correct (n, %)		No ID (n,%)		Mis ID (n, %)		Total	Correct (n, %)		No ID (n, %)		Mis ID (n, %)		Total	Correct (n, %)		No ID (n, %)		Mis ID (n, %)	
cohnii group	11	11	100	0	0	0	0	7	6	85.7	1	14.3	0	0	1	0	0	1	100	0	0
epidermidis group	34	**33**	97.1	1	2.9	0	0	25	22	88.0	2	8.0	1	4.0	102	69	67.6	23	22.5	10	9.8
haemolyticus group	26	25	96.2	0	0	1	3.8	21	15	71.4	5	23.8	1	4.8	45	27	60	14	31.1	4	8.9
lugdunensis group	7	4	57.1	3	42.9	0	0	1	1	100	0	0	0	0	9	0	0	6	66.7	3	33.3
saprophyticus group	11	9	81.8	1	9.1	1	9.1	8	6	75.0	0	0	2	25.0	2	0	0	2	100	0	0
warneri group	8	7	87.5	0	0	1	12.5	10	4	40	3	30	3	30	11	0	0	9	81.8	2	18.2
Total	**86**	**78**	**90.7**	**5**	**5.8**	**3**	**3.5**	**65**	**48**	**73.8**	**10**	**15.4**	**7**	**10.8**	**170**	**96**	**56.5**	**55**	**32.4**	**19**	**11.2**

Table 3.14 Non-represented Gram-negative genera identification results in the test set

	Total	**Correct ID (n, %)**		**No ID (n, %)**		**Mis-ID (n, %)**	
Brevundimonas	2	0	0	0	0	2	100
Chryseobacterium	8	0	0	2	25	6	75
Cupriavidus	2	0	0	1	50	1	50
Delftia	1	0	0	0	0	1	100
Edwardsiella	1	0	0	1	100	0	0
Kingella	3	0	0	0	0	3	100
Moraxella	1	0	0	1	100	0	0
Morganella	12	0	0	5	41.7	7	58.3
Ochrobactrum	7	0	0	4	57.1	3	42.9
Pantoea	10	0	0	7	70	3	30
Paracoccus	2	0	0	2	100	0	0
Pasteurella	1	0	0	0	0	1	100
Plesiomonas	4	0	0	2	50	2	50
Prevotella	1	0	0	0	0	1	100
Providencia	3		0	2	66.7	1	33.3
Raoultella	4	0	0	3	75	1	25
Sphingomonas	2	0	0	0	0	2	100
Yersinia	4		0	3	75	1	25
TOTAL	**68**	**0**	**0**	**33**	**0.49**	**35**	**0.51**

Figure 3.9 PCA of transflection FTIR spectra of Corynebacterium isolates, confirming spectral difference between outlier (red) and correctly identified isolates (blue).



3.4.2.1 Limitations of transflection FTIR spectral-based microbial identification system

Although transflection FTIR spectroscopy has the potential to be implemented in laboratories as a routine microbial identification method, there are a few limitations to the technique that must be addressed. In this study, main issues in the spectral database arose in three parts: a) misidentification of *Bacillus* isolates, b) lack of identification for some of the represented genera and species, in particular, the genera within Enterobacteriaceae family, and the species identification of CoNS isolates, and c) misidentification of isolates that are not well represented, or represented at all in the spectral database. There are potential explanations that address these results; 1) the lack of spectral representation in database, 2) high spectral variability within the categorized group that could not be accounted for in the database, and 3) lack of spectral variability between groups of interest. While some genera were spectrally unique enough to correctly identify them with less than 10 isolates present in the database, such as *Corynebacterium* and *Micrococcus*, (9 and 5 isolates respectively), some genera and species had poor identification results, despite having many more isolates represented in the database. This indicates that the minimum required isolate (and spectral) representation in the database is genus and species dependent.

3.4.2.1.1 Effect of increased isolate representation in the spectral database

One way to determine if identification was to increase the representation of genera and species in the spectral database. Although the proportions of all genera and species in the database could not be equal, an updated database with increased numbers of isolates per genera/species was developed, With the remaining isolates that were not added to the database, the identification results were determined and compared to the original results. Isolates of all genera and species with the exception of *Achromobacter*, *Escherichia*, *Proteus* and *Salmonella,* and *E. faecalis, E. faecium* and *S. aureus* were added to the database. Representation of *E. faecalis*, *E. faecium*, *S. aureus*, and *Salmonella* species were decreased in the updated database, to offer a more balanced proportion in representation. One of the concern was an over representation of certain species such as *E. faecalis, E. faecium* and *S. aureus,* compared to genera such as *Serratia*, *Citrobacter* and *Enterobacter* which represented 0.2 % (*n* = 8), 5.3 % (*n* = 21) and 5.6 % (*n* = 22) of the GN bacteria spectral database respectively. *Salmonella* isolates alone made up 24.7 % (*n* = 97) of the spectra representing GN bacteria in the spectral database. Due to the low number of isolates of *Serratia marcescens* (*n* = 16) and *E. gallinarum* (*n* = 18) available, they were removed from the spectral database.

The adjustment of genus and species representation in the spectral database showcased that the with a more balanced representation, identification can be improved. Detailed results from the updated spectral database are available in the Appendix (Supplementary Table 1 - Supplementary Table 4).The discrimination of isolates in the training sets were improved for several GN genera, such as *Aeromonas* (from 64.3 % to 78.9 %), *Burkholderia* (from 84.6 % to 90 %), *Citrobacter* (from 57.1% to 69.2%), *Enterobacter* (from 72.7% to 80.0%), *Escherichia* (from 89.8 % to 91.5 %), *Klebsiella* (from 95.1 % to 97.9 %), *Pseudomonas* (from 84.8 % to 87.5 %), *Shigella* (from 96 % to 96.7 %), and *Stenotrophomonas* (from 80.0 % to 90.0 %). Despite no changes in the number of isolates in training, *Escherichia* isolates in the training set improved in identification from other Enterobacteriaceae genera, most likely as a result of increased representation of the other genera. *Salmonella* was the only GN genus that had decreased discrimination performance in the updated database (90.7 % to 83.3 %) that may be attributed to the reduced representation from 97 to 60 isolates. The decreased performance for *Salmonella* identification was confirmed through lower correct identification rates in validation and test set combined (80.8 %, *n* = 59 in original spectral database 75.5 % *n* = 83 in updated spectral database). All the isolates that were

initially incorrectly identified or not identified had the same identification results using the updated database. The original database had 9 isolates (9.3 %) that could not be identified at the GN genus level in the training set, and these isolates remained as outliers in the training (n = 7) and test (n = 2) set of the updated database. Despite having a training set that improved in discrimination, *Escherichia* isolates in the validation and test set performed worse in the updated database, with increase in isolates that could not be identified past GN (n = 18 to n = 22). *Enterobacter* isolates had overall improved identification in both training and validation/test set after increasing representation in the database from 22 to 30 isolates.

All CoNS subgroups had improved identification rates in the validation/test sets after increasing their represented isolates in the training set except lugdunensis group and cohnii group which had decreased and the same correct identification performances respectively. All CoNS subgroups except lugdunensis group performed worse in terms of correct identification rates in the training set in the updated database. However, the improved identification at this level for the validation and test sets combined show that the updated database was more robust and able to more reliably predict isolates to the correct subgroup within the CoNS. The results show that with further addition of isolates into the database to represent various CoNS species would improve the reliability and robustness of the transflection FTIR spectral database. The number of isolates that were not-identifiable past CoNS and misidentified at the subgroup level decreased with the updated database, resulting in 36 and 16 isolates less respectively for not-identified and misidentified results. Warneri group did not have improved identification results in the validation/test set, as only two isolates were consistently correctly identified.

There were 6 misidentified isolates based on the provided identification results, that were 3 pairs of isolates had two isolates originating from the same sample, indicated by a dashed number (i.e. -1, -2, -3, etc.). These dashed numbers indicate that multiple colonies of different morphology were isolated from the same sample and identified. These 6 incorrectly identified isolates had been predicted with high confidence as the corresponding other dashed number isolate identification from the same sample (Supplementary Table 5). For example, sample XVIII291-2 was identified as *Acinetobacter*, despite being identified as *Klebsiella pneumoniae* by MALDI-TOF MS, the reference method. However, its counterpart sample, XVIII291-3 (original identification *Acinetobacter baumanii complex*) was identified as *Klebsiella* by the spectral database both in the original and updated versions of the database. Both of these isolates were predicted by the spectral

database with high probability over 97 %. The results did not change after the update in the database. Likewise, across the training, validation and test sets, there were 198 isolates (7.8 %) that had predicted the same results incorrectly (or a lack of identification at the genus or species level) using the original and updated databases. At this time, isolates could not be reanalyzed for confirmation.

One limitation in this study was that there were insufficient number of isolates for many genera evaluated, making it difficult to conclude whether the database truly improved or not. For example, *Klebsiella* had improved discrimination in the training set, after increasing the number of isolates representing the genus from 41 to 48. However when looking at the validation and test sets, there were 4 isolates that were incorrectly identified as another GN genus, and another 4 that were not furthered identified past Enterobacteriaceae by both the original and updated spectral databases. Only one isolate that was originally identified as Enterobacteriaceae was correctly identified as *Klebsiella* with the updated spectral database When observing each isolate's results, the incremental improvement in the identification is noticeable. Yet, when reporting in percentage, the percent correct in the validation and test sets combined decrease from 79.1 % to 77.8 %, as a result of reduced overall tested isolates (n = 43 to n = 36). Another example is *Citrobacter* isolates. There were only 33 isolates total, and therefore when increasing the number of isolates in the spectral database, the test/validation set decreased from 12 to 7 isolates, making it difficult to determine if the addition of 5 isolates in the training set improved the *Citrobacter* identification capabilities of the database. In the training set, the *Citrobacter* identification improved with increased isolates, where 3 isolate that originally could not be identified past GN was correctly identified. There was only one isolate that correctly identified as *Citrobacter* in the validation and training set in both original and updated database. The updated spectral database had 3 misidentified isolates, compared to 2 in the original spectral database. Both *Shigella flexneri* and *S. sonnei* had decreased identification performance with the updated spectral database. However, the incorrect isolates were consistent between the two databases, and the decrease in correct identification rate was due to the smaller number of isolates tested in validation/test, as a result of increasing representation in the training set.

Overall, the addition of isolates into training had improved the identification results. This should be further evaluated when more spectra acquired from new samples are available to test the performance of the spectral database. A key thing to note also from the updated database, was that

the decrease in representation of *S. aureus, E. faecalis* and *E. faecium* did not affect the capability of the spectral database to identify these three species correctly. All three isolates had over 98 % correct identification rates in the validation/test sets, with $n = 1022$, $n = 85$ and $n = 519$ respectively.

3.4.2.1.2 Misidentification of *Bacillus* isolates: A case of high spectral and biochemical differences within the genera

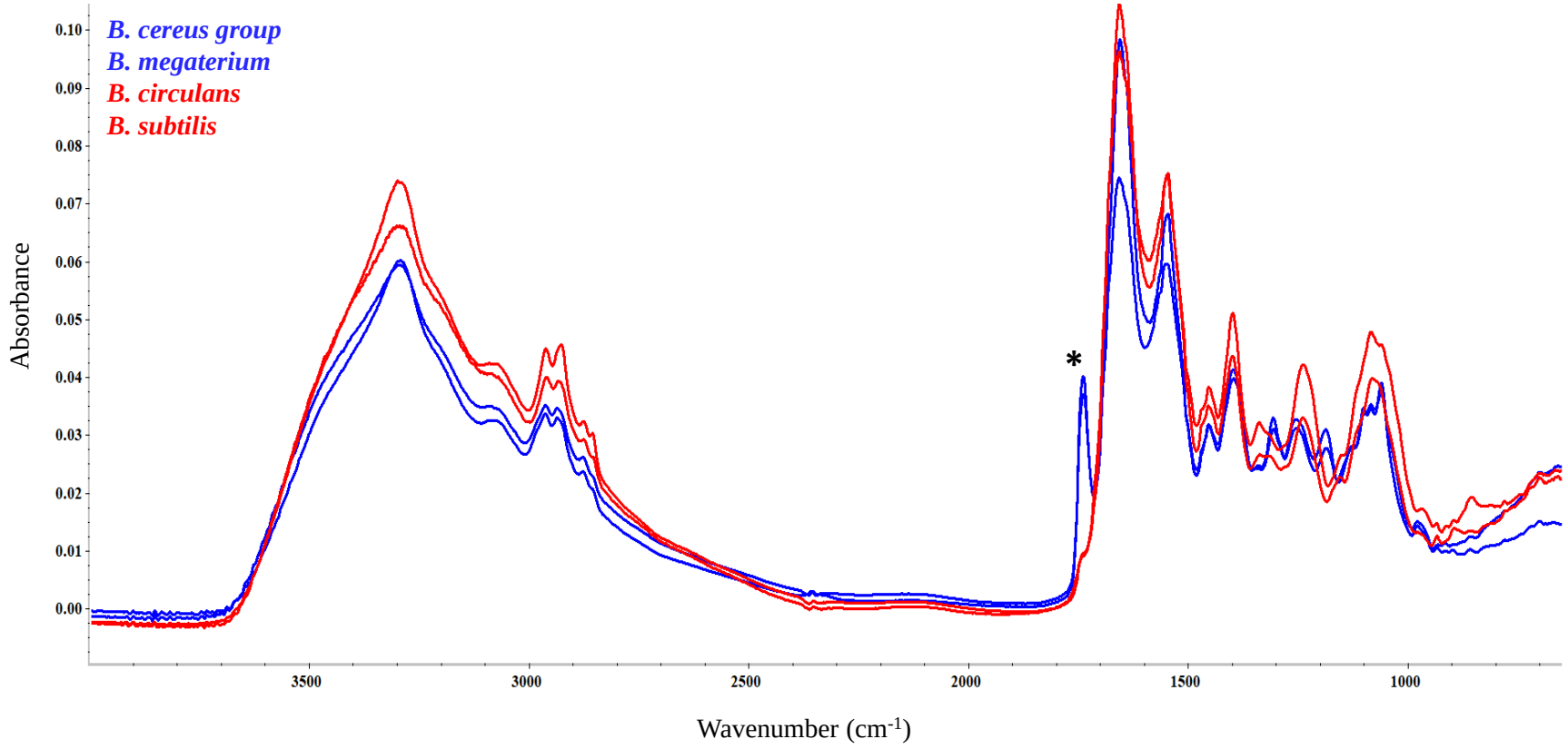
Bacillus isolates had low correct identification based on the developed spectral database. Errors occurred at multiple levels of the database, starting off with incorrect identification at the Gram-level, followed by misidentification at the genus level. *Bacillus* species are known to be diverse in their phenotype, with some species being Gram variable [39, 40]. The variability in *Bacillus* species morphology and phenotypic characteristics was also reflected in the FTIR spectra (Figure 3.10). Three *Bacillus* isolates in the test set were incorrectly predicted at the Gram-level, identified as Gram-negative isolates. The three-dimensional PCA plot (Figure 3.11) based on data points in the spectral region 1350-800 cm^{-1}, displayed how relative to spectra of Gram-positive isolates (green) and Gram-negative isolates (red), the spectra of the three *Bacillus* outlier isolates (purple) are not similar to either groups. However, due to the nature of binary SVM analyses, spectra were required to be predicted as one of the two classes. This resulted in the isolates predicting as Gram-negative bacteria, since it was spectrally more similar to the spectra of Gram-negative bacteria than to Gram-positive bacteria in the database. These outlier isolates were not identified to the species level by reference methods, and therefore, it was not possible to further analyze and determine whether the GN identification by the FTIR spectral database reflected the Gram-variable nature of these *Bacillus* isolates [39].

Similarly to the Gram-negative identification, some *Bacillus* species were spectrally more similar to *Staphylococcus* species, rather than to other *Bacillus* isolates (Figure 3.12) The difference was visible to the naked eye in the raw FTIR spectra when comparing *Bacillus cereus group*/*B. megaterium* and *B. subtilis*/*B. circulans* at peaks around 1740 cm^{-1}, related to the C=O stretching of carbonyl bands (Figure 3.10). This is in concordance with the groupings of *Bacillus* species based on morphology and recent taxonomic development, which splits the *Bacillus* genus into two groups, the *B. subtilis* group, which includes *B. circulans, B. coagulans, B. lichenformis, B. pumilus* and *B. subtilis*, and the *B. cereus* group, including species such as *B. anthracis, B. cereus, B. megaterium, B. mycoides* and *B. thuringiensis* [39]. *B. circulans* and *B. subtilis* spectra

have more spectral features that are similar to *Staphylococcus* (Figure 3.12). This was further demonstrated in Figure 3.13 where the HCA (cosine distance metric and ward linkage type) in the region 1350-800 cm^{-1} showcased spectral similarity of *B. circulans* and *B. subtilis* to *S. aureus* rather than to *B. cereus group* isolates. As per Figure 3.10 and Figure 3.12, the spectra of one of the incorrectly identified *Bacillus species* isolate is likely to be an isolate that is not part of the *Bacillus cereus* group. Despite understanding that there are spectral similarities between *Staphylococcus species* and some *Bacillus species*, the lack of *B. subtilis* group isolates available during data acquisition ($n = 1$) prevented the development of an additional analysis level that allows for *Bacillus species* identification. Furthermore, 12 isolates (34.3%) were only identified as *Bacillus species* by the reference method(s), thereby hindering the use of these isolates for species specific classification and identification. By adding more *Bacillus* isolates that have been identified to the species level into the database, it may be possible to predict *Bacillus* isolates correctly.

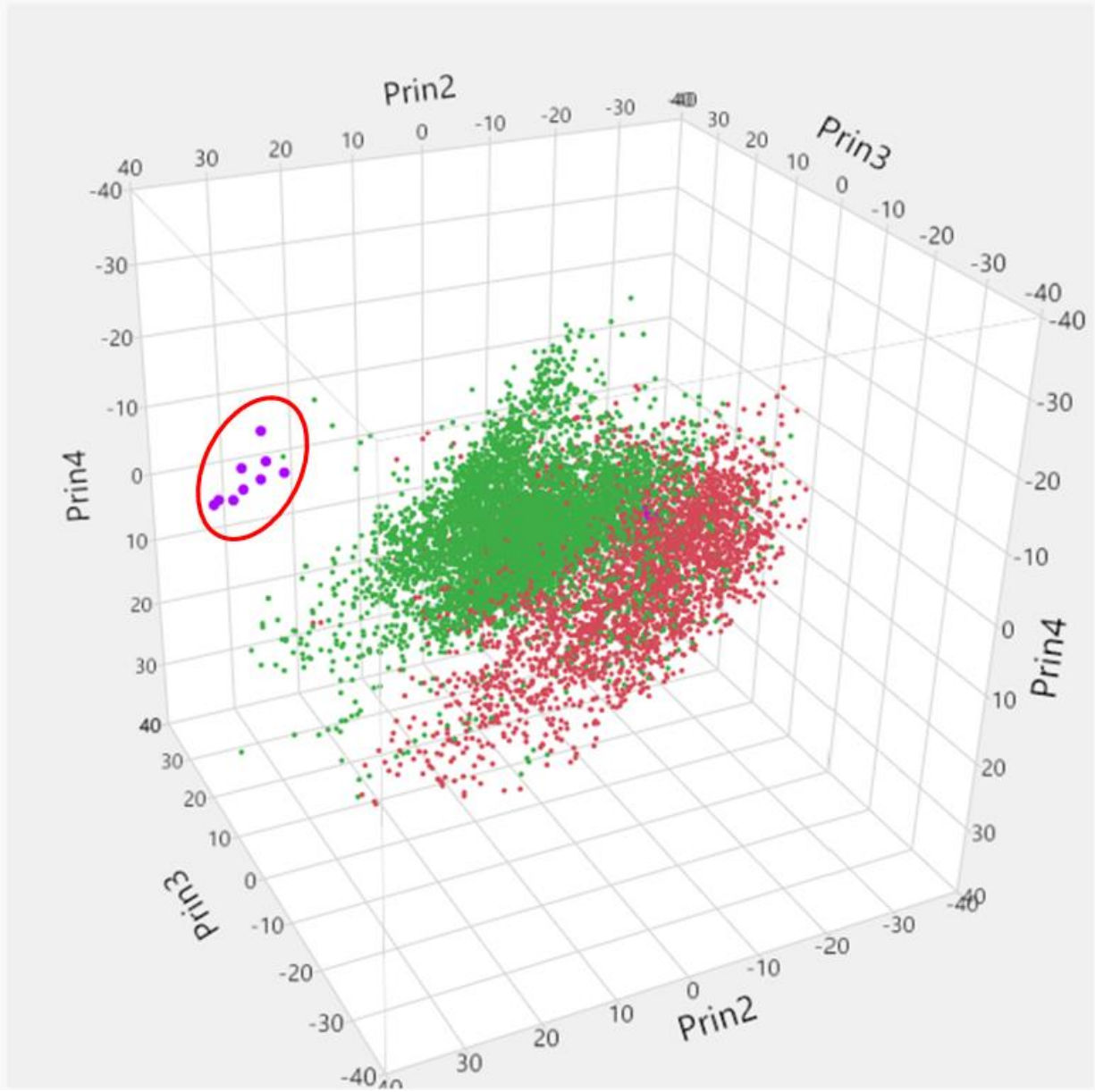
Spectral differences between averaged FTIR spectra of GP genera and species that were present in the spectral database were showcased, both in broad region (1350-800 cm^{-1}) and with region selection to maximize difference between the genera/species (Figure 3.14). GP genera showcased that with the exception of *Bacillus* species, spectral differences between the genera were prominent.

Figure 3.10 Overlay of raw transflection FTIR spectra of Bacillus *species*



Raw average FTIR spectra of *B. cereus* group and *B. megaterium* isolates in blue, overlayed with raw average FTIR spectra of *B. circulans* and *B. subtilis* isolates (red). Overlay spectra show differences in features, notably the 1738 cm^{-1}, related to lipids [41]. 1745 cm^{-1} C=O Stretching attributed to lactam of muramic acid, or saturate esters (covalent link between teichoic and teichuronic acid) in *Bacillus subtilis* [42].

Figure 3.11 Three-dimensional principal component plot visualizing the spectral dissimilarity of FTIR spectra acquired from Bacillus *isolates (purple dots, circled in red) to both Gram-negative (red dots) and Gram-positive (green dots)*



Three *Bacillus* isolates that were incorrectly predicted at the Gram-level observed unique features in the spectral region 1350-800 cm^{-1}. The three-dimensional PCA plot shows how relative to spectra of Gram-positive bacterial isolates (green) and Gram-negative bacterial isolates (red), the spectra of the three Bacillus isolates are not similar to either groups.

Figure 3.12 Overlay of raw transflection FTIR spectra of Bacillus circulans *(red on top figure,),* Bacillus subtilis *(red on top figure) and* Staphylococcus aureus *(green). Spectrum of outlier isolate (red, bottom figure) showcasing spectral similarity to* S. aureus *rather than* Bacillus cereus *and* Bacillus *spectra (blue spectra, bottom figure)*

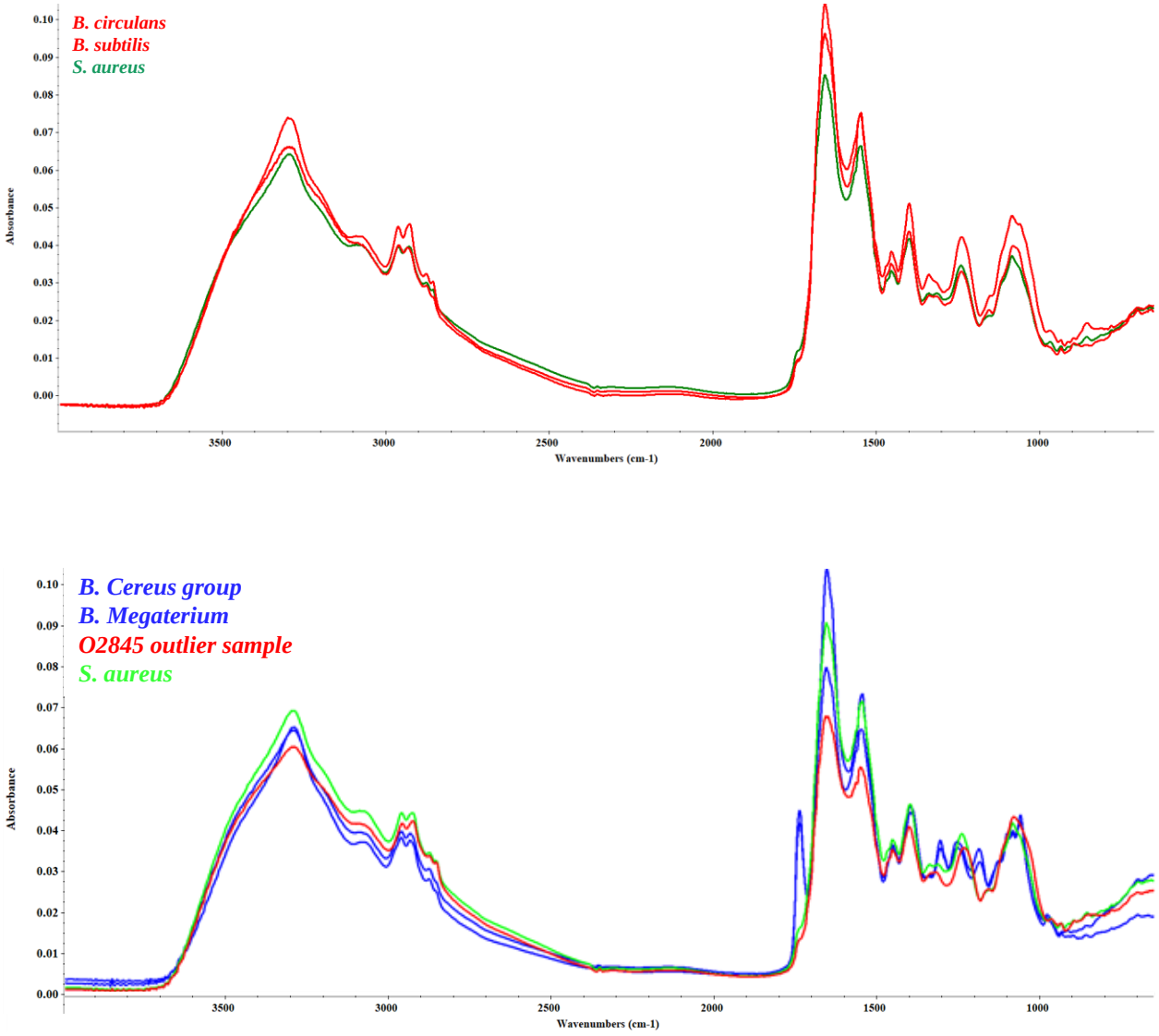


Figure 3.13 Bacillus *species spectral difference based on spectral region 1350-800 cm^{-1} visulaized in a dendrogram generated by HCA indicate spectral similarity of* S. aureus *(blue),* B. circulans *(Orange) and* B. subtilis *(Purple).*

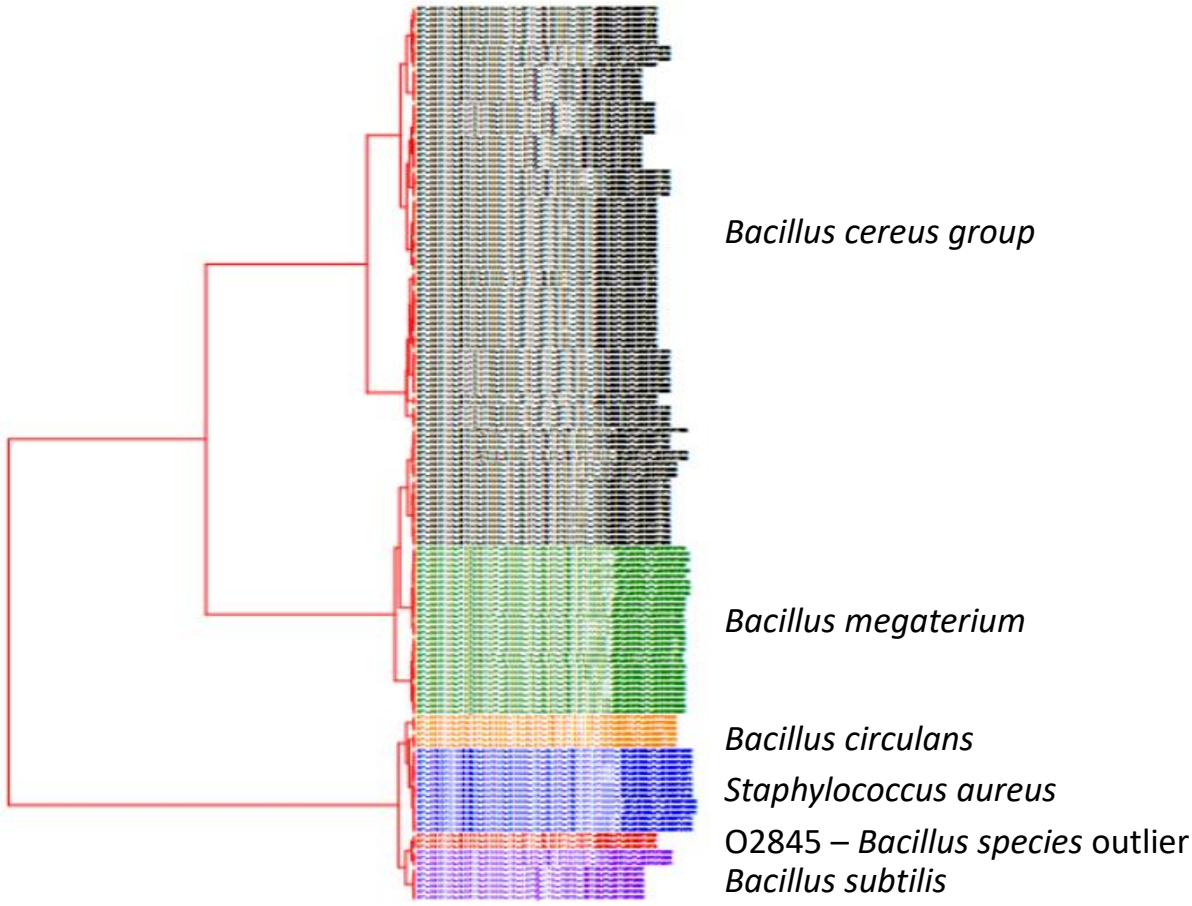
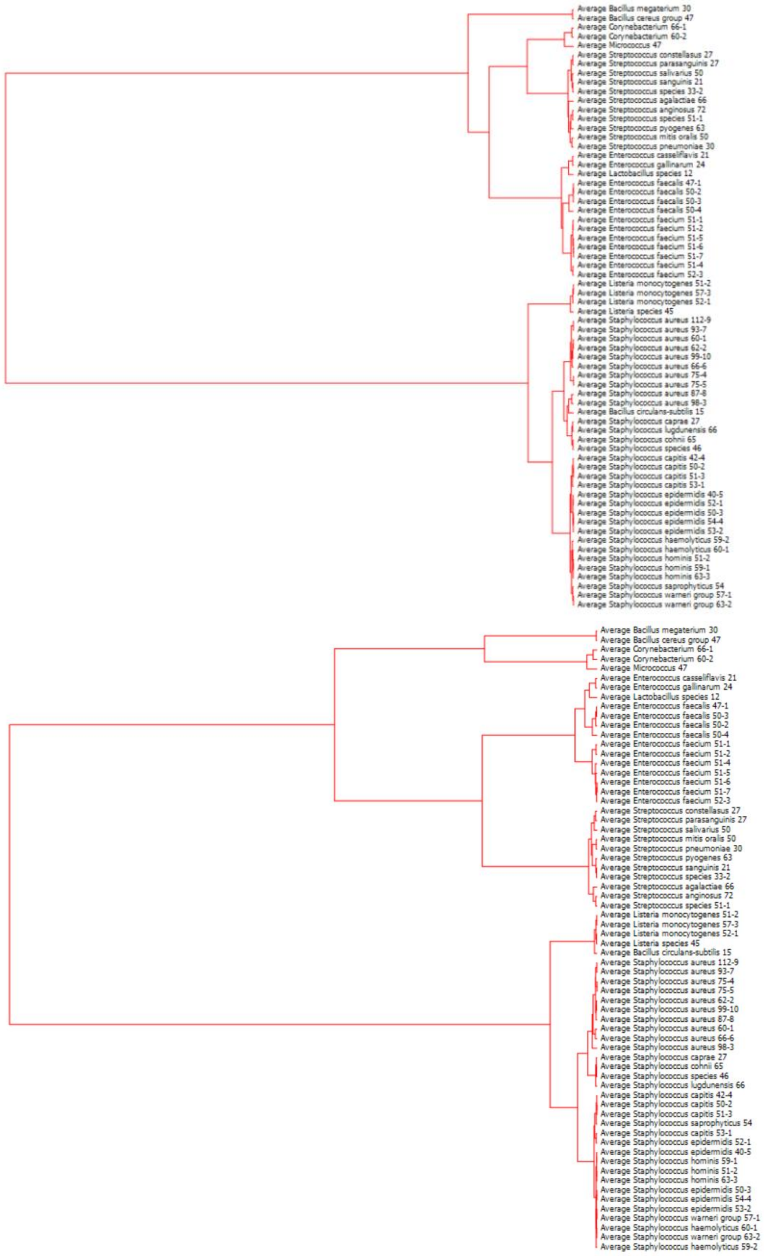


Figure 3.14 HCA of averaged FTIR spectra of GP genera in selected regions (without (left) and with (right) the use of forward search) between spectral region 1350-800 cm^{-1}, showcasing the relative similarities and dissimilarities that enable for successful GP genera identification by FTIR spectroscopy.



Averaged spectra of *Bacillus cereus* and *B. megaterium* are not similar to *Bacillus circulans*-subtilis average spectrum, which is clustering closest to the averaged *Listeria* spectra.

3.4.2.1.3 Misidentification of Gram-negative bacterial isolates at the genus level

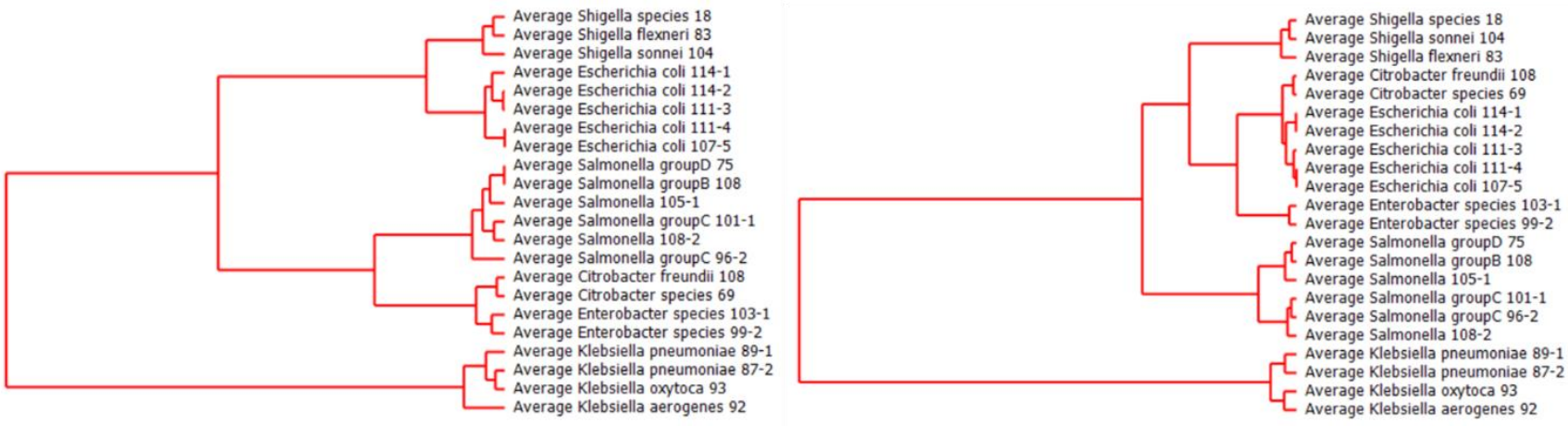
There were 7 Enterobacteriaceae genera in the transflection FTIR spectral database. At the genus level, concordant classification in the training set, relative to reference methods ranged from 57.1 % to 96 %, where *Citrobacter* and *Enterobacter* performed particularly poorly. The concordance between the FTIR spectral database and reference methods for identification in the validation and test sets were lower than the training set, which indicates poor modelling in the training set, potentially as a result of overfitting. This portion of the database proved to be unreliable for robust identification. Similarly to the spectral analyses conducted for GP genera and species, averaged GN genera and species spectra were visualized in HCA to showcase their relative spectral similarities in both broad and specific regions (identified by forward region search). In Figure 3.15, HCA showed that *Klebsiella* species are spectrally unique from other Enterobacteriaceae genera, with selected spectral regions from 1350-800 cm^{-1} with (left figure) and without (right figure) the 2800-3000 cm^{-1}, which corresponds to the C-H absorption, strongly associated with lipids [43]. With spectral region selection, there were spectral regions that enabled discrimination amongst the Enterobacteriaceae genera. It is interesting to note that the addition of the C-H region (2800-3000 cm^{-1}) increased spectral difference between spectra of *E. coli* and *Citrobacter* species, while spectral difference between *Shigella* and *E. coli* decreased. Additionally, the spectral difference between *Citrobacter* and *Enterobacter* decreased with the added C-H region. *Citrobacter* has great phenotypic variability as well., with atypical strains of *C. freundii* frequently identified. It was reported previously that there are strain types that are more similar to *Salmonella* and *Escherichia coli* [44, 45]. *Citrobacter* is reported to cause false positive results for *Salmonella* detection in foods, due to similarities in the metabolic and antigenic properties [46]. Therefore, it is not surprising that the spectral similarities between these Enterobacteriaceae genera are high, and resulted in high misidentification rates.

Additionally, in Figure 3.16, spectral difference between all GN genera that were part of the database were visualized in HCA using spectral regions identified from forward search algorithm. Although most non-Enterobacteriaceae genera were clustered away from Enterobacteriaceae genera, there were some exceptions. First, *Aeromonas* spectra were found in a cluster closest to spectra of genera such as *Shigella* and *E. coli*. Based on broad spectral regions, 1800-800 cm^{-1}, and 2800-3000 cm^{-1}, *Aeromonas* spectra were most similar to *Salmonella* spectra in HCA (data not shown). *Aeromonas* is known to have many species that are genotypically and

phenotypically diverse. Many of its species, including *A. hydrophila, A. veronii* and *A. caviae,* which are species that were represented in the database, are known to have open pangenomes, and contain high proportions of accessory genes (62 %, 53 %, 51 % respectively). This is known to be related to high variation in virulence genes and antimicrobial resistance profiles [47]. Additionally, it is known that identification based on phenotype is difficult for *Aeromonas* isolates, as it is recognized to be phenotypically similar to *Vibrio*, and result in low concordance to phylogenetic identification methods such as *rpoB* sequencing [48]. It is likely that *Aeromonas* representation in a database constructed with 14 *Aeromonas* isolates was underrepresented in trying to achieve robust identification. Despite most errors that occurred in for *Aeromonas* isolates consisted of identification as Enterobacteriaceae or *Salmonella*, no reports have suggested that there are biochemical or genotypic similarities between *Aeromonas* and *Salmonella.*

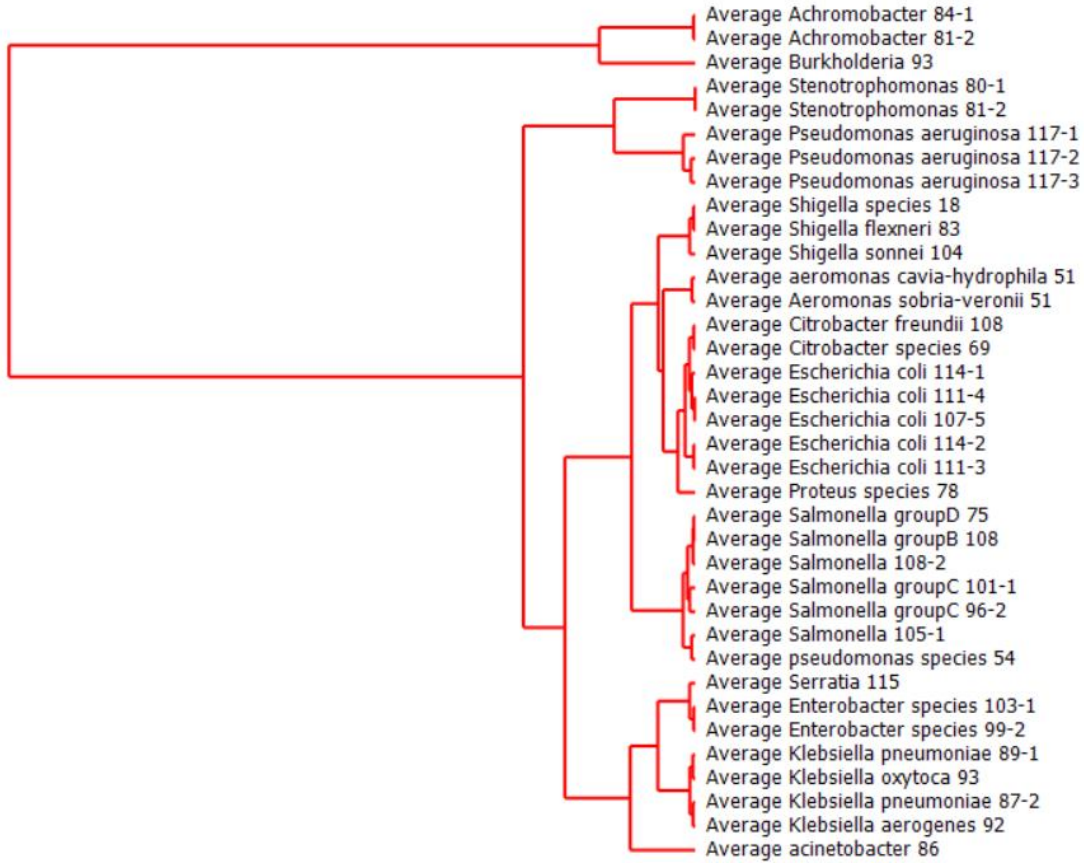
Another genus that showed spectral similarity to Enterobacteriaceae was non-aeruginosa *Pseudomonas* species. Although in the database structure, discrimination between Enterobacteriaceae and non-Enterobacteriaceae genera was built-in as an intermediate step for GN genera identification, based on spectral region selection it is now apparent that discriminating bacterial isolates based on family classification may not be suitable for some genera. Furthermore, for genera like *Serratia* and *Acinetobacter,* which were not part of the region selection algorithm, despite not being a part of the Enterobacteriaceae family, they were observed to be more spectrally similar to them than to non-Enterobacteriaceae bacteria.

Figure 3.15 HCA of averaged Enterobacteriaceae genera with spectral regions identified by forward search algorithm in regions 1350-800 cm^{-1} and 2800-3000 cm^{-1} (left) and 1350-800 cm^{-1} (right).



Enterobacteriaceae genera have spectral differences that should enable discrimination between them. The discrimination was not as successful when doing region selection in 1350-800 cm^{-1}, without the lipid regions 2800-3000 cm^{-1}. This indicates that the lipid regions are more critical in discrimination between Gram-negative bacteria. This coincides with the fact that Gram-negative bacteria have a lipid layer as part of the cell membrane/wall. Without the use of the lipid regions, the discrimination between *Citrobacter* and *Escherichia* are less obvious. *Klebsiella* and *Shigella* isolates appear to be the most spectrally distinct from other Enterobacteriaceae genera

Figure 3.16 HCA of averaged GN genera with regions identified by forward search algorithm in regions 1350-800 cm^{-1} and 2800-3000 cm^{-1}



Although the HCA includes *Serratia, Acinetobacter, Burkholderia,* and *Proteus*, these genera were not included in the region selection process. While some genera such as *Achromobacter, Burkholderia, Pseudomonas (aeruginosa)* and *Stenotrophomonas* are spectrally unique and do not cluster with Enterobacteriaceae genera isolates, genera such as *Proteus, Serratia* and *Aeromonas* clustered within the Enterobacteriaceae genera.

3.4.2.1.4 Misidentification of isolates that are underrepresented in the database

Although the spectral database was developed using genera and species that are commonly associated with clinical and food-industry settings, there are many microorganisms that must be identified when identifying the causative agent of a patient's infection. Because these rarely occurring or newly emerging microorganisms are difficult to encounter, the number of isolates that were available for data collection in this experiment was small. Lactic acid bacteria such as *Lactobacillus* and *Lactococcus* isolates were misidentified with medium to high confidence level as *Enterococcus* species. These errors, that result from high biochemical and spectral similarity can only be resolved through the addition of the underrepresented genera. It is not surprising at all that isolates of these genera were misidentified as *Enterococcus* species, as they are all part of the lactic acid bacteria group. Successful discrimination between *Enterococcus* species and other lactic acid bacteria have been demonstrated previously, indicating that with sufficient representation of these genera, the database can also successfully identify *Lactobacillus* and *Lactococcus* isolates, thereby reducing false positive identification of *Enterococcus* species [35]The lack of representation in the spectral database is a problem which can be addressed relatively easily, and improved upon over time through the spectral addition of well-characterized isolates of more (new and rare) genera and species. Emerging and rare microorganisms can increasingly be identified as databases are updated periodically. Other techniques such as MALDI-TOF MS, which also rely on the use of reference databases for identification, also run into this problem. Improved identification of clinically relevant yeast species using ATR-FTIR spectroscopy-based technique was previously demonstrated through the expansion of the database [13]. In the current database, only 7 Gram-positive, 14 Gram-negative and 1 yeast genera were represented, limiting the use for rare microorganisms in the clinical and food-industry settings.

3.5 Conclusion

In this study, transflection FTIR spectral acquisition mode was first compared in sample preparation, spectral acquisition and spectral quality relative to ATR and transmission FTIR spectral modes, which are the two commonly used spectral acquisition modes for microbial studies. Transflection spectral acquisition, although not as popular as ATR and transmission modes, showcased high spectral quality and discriminatory capabilities that were comparable to ATR and transmission modes, suitable for microbial identification and discrimination purposes. This was further confirmed through the development of a transflection-FTIR spectral database, using spectra acquired from a limited number of genera and species of Gram-positive and Gram-negative bacteria and yeast isolates that are relevant in clinical and food microbiology. The scope of this study was to evaluate whether developing a spectral database using transflection FTIR data was possible, and to determine possible limitations in this approach. Based on the current study, a spectral database developed using a series of PCA and SVM was able to achieve highly concordant identification for species such as *Staphylococcus aureus*, *Enterococcus faecalis* and *Enterococcus faecium* relative to reference methods such as VITEK 2 (biochemical assays), and MALDI-TOF MS. The database also showed that with lacking representation in the database, the confidence level as well as the number of correctly identified isolates deteriorate and become less reliable. However, with increased representation in the spectral database, identification accuracy and robustness can be improved.

3.6 References

1. 1. Váradi, L., et al., *Methods for the detection and identification of pathogenic bacteria: past, present, and future.* Chemical Society Reviews, 2017. **46**(16): p. 4818-4832.
2. Seng, P., et al., *MALDI-TOF-mass spectrometry applications in clinical microbiology.* Future microbiology, 2010. **5**(11): p. 1733-1754.
3. Lévesque, S., et al., *A side by side comparison of Bruker Biotyper and VITEK MS: utility of MALDI-TOF MS technology for microorganism identification in a public health reference laboratory.* PloS one, 2015. **10**(12): p. e0144878.
4. Kostrzewa, M. and S. Schubert, *MALDI-TOF mass spectrometry in microbiology*. 2016: Caister Academic Press.
5. Wenning, M. and S. Scherer, *Identification of microorganisms by FTIR spectroscopy: perspectives and limitations of the method.* Applied microbiology and biotechnology, 2013. **97**(16): p. 7111-7120.
6. Novais, Â., et al., *Fourier transform infrared spectroscopy: unlocking fundamentals and prospects for bacterial strain typing.* European Journal of Clinical Microbiology & Infectious Diseases, 2019. **38**(3): p. 427-448.
7. Kirkwood, J., et al., *Differentiation of group I and group II strains of Clostridium botulinum by focal plane array Fourier transform infrared spectroscopy.* Journal of food protection, 2006. **69**(10): p. 2377-2383.
8. Ghetler, A., *Development of an expert system for the identification of bacteria by focal plane array Fourier transform infrared spectroscopy*. 2010, McGill University.
9. Carranza, L., et al., *EVALUATION OF FPA-FTIR SPECTROSCOPY AS A TOOL IN THE DIFFERENTIATION OF CAMPYLOBACTER JEJUNI FROM CAMPYLOBACTER COLI ISOLATED FROM RETAIL CHICKEN SAMPLES.* Journal of Food Safety, 2012. **32**(3): p. 289-295.
10. Baker, M.J., et al., *Using Fourier transform IR spectroscopy to analyze biological materials.* Nature protocols, 2014. **9**(8): p. 1771.
11. Wehbe, K., et al., *The effect of optical substrates on micro-FTIR analysis of single mammalian cells.* Analytical and bioanalytical chemistry, 2013. **405**(4): p. 1311-1324.
12. Nitrosetein, T., et al., *Attenuated Total Reflection Fourier Transform Infrared Spectroscopy combined with chemometric modelling for the classification of clinically relevant Enterococci.* Journal of Applied Microbiology, 2020.
13. Lam, L.M., et al., *Reagent-Free Identification of Clinical Yeasts Using Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy.* Journal of clinical microbiology, 2019: p. JCM. 01739-18.
14. Wohlmeister, D., et al., *Differentiation of Candida albicans, Candida glabrata, and Candida krusei by FT-IR and chemometrics by CHROMagar™ Candida.* Journal of microbiological methods, 2017. **141**: p. 121-125.
15. DeSouza, L., T. Kamat, and C. Naik, *Diffuse reflectance infrared fourier transform spectroscopic (DRIFTS) investigation of E. coli, Staphylococcus aureus and Candida albicans.* 2009.
16. Timmins, É.M., et al., *Rapid differentiation of closely relatedCandida species and strains by pyrolysis-mass spectrometry and fourier transform-infrared spectroscopy.* Journal of Clinical Microbiology, 1998. **36**(2): p. 367-374.

17. Goodacre, R., et al., *Rapid identification of Streptococcus and Enterococcus species using diffuse reflectance-absorbance Fourier transform infrared spectroscopy and artificial neural networks.* FEMS Microbiology Letters, 1996. **140**(2-3): p. 233-239.
18. DeVetter, B.M., et al., *Characterization of the structure of low-e substrates and consequences for IR transflection measurements.* Vibrational Spectroscopy, 2017. **91**: p. 119-127.
19. Sedman, J., et al., *Infrared Imaging: principles and practices.* Handbook of Vibrational Spectroscopy, 2010.
20. Lam, T.M.L., *Rapid Identification and Classification of Escherichia Coli and Shigella by Attenuated Total Reflectance-Fourier Transform Infrared Spectroscopy*. 2017, McGill University Libraries.
21. Sousa, C., et al., *Development of a FTIR-ATR based model for typing clinically relevant Acinetobacter baumannii clones belonging to ST98, ST103, ST208 and ST218.* Journal of Photochemistry and Photobiology B: Biology, 2014. **133**: p. 108-114.
22. Bağcıoğlu, M., et al., *Detection and identification of Bacillus cereus, Bacillus cytotoxicus, Bacillus thuringiensis, Bacillus mycoides and Bacillus weihenstephanensis via machine learning based FTIR Spectroscopy.* Frontiers in microbiology, 2019. **10**: p. 902.
23. Grunert, T., et al., *Rapid and reliable identification of Staphylococcus aureus capsular serotypes by means of artificial neural network-assisted Fourier transform infrared spectroscopy.* Journal of clinical microbiology, 2013. **51**(7): p. 2261-2266.
24. Amiali, N.M., et al., *Rapid identification of community-associated methicillin-resistant Staphylococcus aureus by Fourier transform infrared spectroscopy.* Diagnostic microbiology and infectious disease, 2011. **70**(2): p. 157-166.
25. Morais, C.L., F.S. Costa, and K.M. Lima, *Variable selection with a support vector machine for discriminating Cryptococcus fungal species based on ATR-FTIR spectroscopy.* Analytical Methods, 2017. **9**(20): p. 2964-2970.
26. Guliev, R.R., et al., *Discrimination of Staphylococcus aureus Strains from Coagulase-Negative Staphylococci and Other Pathogens by Fourier Transform Infrared Spectroscopy.* Analytical Chemistry, 2020. **92**(7): p. 4943-4948.
27. Flournoy, D. *Facilitating identification of lactose-fermenting enterobacteriaceae on macconkey agar*. in *Proceedings of the Oklahoma Academy of Science*. 1990.
28. Rogalski, A., *Infrared detectors: status and trends.* Progress in quantum electronics, 2003. **27**(2-3): p. 59-210.
29. Wenzel, T., *Molecular and Atomic Spectroscopy Fourier-Transform Infrared Spectroscopy (FT-IR)*. 2020.
30. Enfield, A.L., *Investigation of the High-Throughput Analytical Performance of an FPA-FTIR Imaging System*. 2010, McGill University Library.
31. Amiali, N.M., et al., *Rapid identification of coagulase-negative staphylococci by Fourier transform infrared spectroscopy.* Journal of microbiological methods, 2007. **68**(2): p. 236-242.
32. Davis, J.T., E. Foltz, and W.S. Blakemore, *Serratia marcescens: a pathogen of increasing clinical importance.* JAMA, 1970. **214**(12): p. 2190-2192.
33. Hejazi, A. and F. Falkiner, *Serratia marcescens.* Journal of medical microbiology, 1997. **46**(11): p. 903-912.

34. Farber, J. and P. Peterkin, *Listeria monocytogenes, a food-borne pathogen.* Microbiology and Molecular Biology Reviews, 1991. **55**(3): p. 476-511.
35. Levasseur-Garcia, C., C. Couderc, and H. Tormo, *Discrimination of lactic acid bacteria Enterococcus and Lactococcus by infrared spectroscopy and multivariate techniques.* Journal of Near Infrared Spectroscopy, 2017. **25**(4): p. 231-241.
36. Becker, K., C. Heilmann, and G. Peters, *Coagulase-negative staphylococci.* Clinical microbiology reviews, 2014. **27**(4): p. 870-926.
37. Organization, W.H., *Antimicrobial resistance: global report on surveillance*. 2014: World Health Organization.
38. Larsh, H.W., *Manual of clinical microbiology*. 1975, JSTOR.
39. England, P.H., *Identification of Bacillus species*, in *UK Standards for Microbiology Investigations*. 2018.
40. Turnbull, P.C., J. Kramer, and J. Melling, *Bacillus.* Manual of clinical microbiology, 1991. **5**: p. 296-303.
41. Balan, V., et al., *Vibrational spectroscopy fingerprinting in medicine: from molecular to clinical practice.* Materials, 2019. **12**(18): p. 2884.
42. Filip, Z., S. Herrmann, and J. Kubat, *FT-IR spectroscopic characteristics of differently cultivated Bacillus subtilis.* Microbiological research, 2004. **159**(3): p. 257-262.
43. Quintelas, C., et al., *An overview of the evolution of infrared spectroscopy applied to bacterial typing.* Biotechnology journal, 2018. **13**(1): p. 1700449.
44. Delgado, G., et al., *Genetic characterization of atypical Citrobacter freundii.* PLoS One, 2013. **8**(9): p. e74120.
45. Bettelheim, K., et al., *Isolation of a Citrobacter freundii strain which carries the Escherichia coli O157 antigen.* Journal of clinical microbiology, 1993. **31**(3): p. 760-761.
46. Pilar, A.V.C., et al., *Similar yet different: phylogenomic analysis to delineate Salmonella and Citrobacter species boundaries.* BMC genomics, 2020. **21**: p. 1-13.
47. Gauthier, J., et al., *Strong genomic and phenotypic heterogeneity in the Aeromonas sobria species complex.* Frontiers in microbiology, 2017. **8**: p. 2434.
48. Fernández-Bravo, A. and M.J. Figueras, *An update on the genus Aeromonas: Taxonomy, epidemiology, and pathogenicity.* Microorganisms, 2020. **8**(1): p. 129.

Connecting Statement

Transflection-FTIR spectral acquisition mode showcased that it has the capability to collect FTIR spectra with sufficient and comparable spectral quality to those acquired by transmission and ATR modes. Furthermore, the spectral database developed using transflection-FTIR spectra showcased that with sufficient representation (number of isolates required in database is genus and species dependent), isolates can be identified with high confidence, robustly. In the next chapter, FTIR spectroscopy was evaluated for its capability to identify MRSA, through the combined use of the spectral database and antibiotic containing media.

Chapter 4. Use of Antibiotic-Containing Agar for Rapid MRSA Identification by FTIR Spectroscopy

4.1 Abstract

Staphylococcus aureus is a leading cause of bacterial infections in humans, ranging from skin, wound, and surgical-site infections to potentially life-threatening invasive endocarditis and bloodstream infections. It is critical to rapidly identify methicillin-susceptible (MSSA) and methicillin-resistant *S. aureus* (MRSA) for appropriate treatment of invasive infections, as well as for detecting new colonization in hospitalized patients. Current MRSA screening method for surveillance purpose utilizes a combination of polymerase chain reaction (PCR) and chromogenic agar, where the techniques detect the MRSA with the presence of the *mecA* gene, relating to methicillin-resistance, and color change on the chromogenic agar. The chromogenic agar is a selective and differential medium that is expensive relative to generic culture medium, due to the addition of specific chromogens that enable visual detection of MRSA. Fourier transform infrared (FTIR) spectroscopy has successfully demonstrated rapid identification between *S. aureus* and coagulase-negative staphylococci (CoNS) isolates through the use of a FTIR spectral database. In this study, we use this spectral database in conjunction with a modified growth media protocol, employing the antibiotic cefoxitin, to inhibit growth of MSSA, for FTIR-spectroscopy based MRSA identification. Modified blood agar plates were made using Columbia agar with 5 % sheep blood supplemented with colistin (10 µg/mL) and nalidixic acid (15 µg/mL) (CNA) to inhibit Gram-negative bacteria and with cefoxitin at concentrations of 4 µg/mL (4FOX) and 8 µg/mL (8FOX). 229 staphylococcal isolates that were previously identified by MALDI-TOF MS, were cultured on BAP, 4FOX-CNA-BAP, 8FOX-CNA-BAP and MRSA selective chromogenic agar. Isolated colonies were deposited onto IR-reflective slides for FTIR spectral acquisition and identified as *S. aureus* or CoNS by matching their spectra against the FTIR spectral database. Isolates were identified as MRSA if colonies growing on cefoxitin-containing agar were spectrally identified as *S. aureus*. All *S. aureus* (n = 99) and CoNS (n = 130) isolates were correctly identified using the FTIR spectral database. The modification in media made by the addition of antibiotics did not affect the identification of *S. aureus* from CoNS. In reference to VITEK 2 AST results, MRSA was identified with 100 % sensitivity and 95.3 % specificity on 4FOX-CNA-BAP after 24 hours of incubation, achieving 98 % categorical agreement. 100 % sensitivity and 91.1 % specificity was achieved after 24 hours incubation when isolates were grown on 8FOX-CNA-BAP.

These results were comparable to those obtained with MRSA selective chromogenic agar (100 % sensitivity and 97.8 % specificity). This FTIR spectroscopy based method for MRSA identification using 4 µg/mL cefoxitin achieved high sensitivity and specificity for MRSA through accurate identification of *S. aureus* isolates on cefoxitin-containing agar. As an MRSA screening tool, this methodology may provide an alternative to the use of costly chromogenic media at substantially lower cost.

4.2 Introduction

Staphylococcus aureus is a commensal microorganism which can become invasive, causing infections in skin, soft tissues, organs, and bloodstreams. It is known to be one of the most common pathogens in both health-care and community associated infections [1]. Accurate and rapid identification of *Staphylococcus aureus*, especially those that are methicillin-resistant (MRSA) is critical in making the appropriate treatment decisions. Resistance to methicillin in *S. aureus* is becoming increasingly prevalent as the use of antibiotics became more frequent in treating both humans and animals. MRSA is listed as one of many antibiotic resistant organisms that are of concern to the World Health Organization (WHO), due to its presence in five of six WHO regions with national reports of 50% resistance or more [2]. MRSA is one of the most prevalent nosocomial pathogens in Canada along with vancomycin resistant enterococci (VRE) and *Clostridium difficile*. In the USA, it is the leading cause of hospital acquired infection (HAI), with approximately 80,000 severe MRSA infections per year, reported in 2013 [3]. Centers for Disease Control and Prevention (CDC) reported an estimated 323,700 cases of MRSA related HAI and 10,600 deaths in 2017. Despite the decrease in overall MRSA infections between 2005 and 2016 (17 % decline per year), decline is suggested to have slowed, with no change between 2013 and 2016 for hospital acquired MRSA bloodstream infections [4]. MRSA colonization and infection are both known to increase morbidity and mortality of patients. Furthermore, infections caused by antibiotic resistant microorganisms result in increased risk of worse clinical outcome and death [2, 3]. Therefore, it is critical that appropriate infection control is implemented and the transmission of MRSA is minimized within health care facilities between the environment, healthcare personnel and patients.

4.2.1 Current identification methods of *Staphylococcus* species and their resistance to antibiotics

Current identification methods for MRSA and MSSA relies on the combined use of phenotypic and genotypic technologies. MALDI-TOF MS is a popular technology in rapid microbial identification due to its ease of sample preparation and use of system [5]. Although it is an excellent and reliable method for identification, it is not approved for detection of antimicrobial resistance. Therefore, additional antibiotic susceptibility testing, such as microbroth dilution or disk diffusion are needed. This requires an additional 16-24 hours to achieve the AST results. Genotypic methods, such as polymerase chain reaction (PCR) is extremely useful in screening, as

it looks for specific target genes that are unique to the organisms of interest. Furthermore, it can be used as confirmation for presence of genes related to antibiotic resistance. PCR identifies the presence of specific gene sequences, such as *nuc* and *mecA* genes, which are used to determine if the sample is *S. aureus* and methicillin resistant respectively [6]. Although multiplex PCR can identify MRSA in a single test, the test is expensive, hindering its use for routine screening, especially in smaller scale laboratories [7]. Furthermore, variants of *mecA* gene, such as *mecC*, which only have 70% homology to *mecA*, makes PCR susceptible to misidentifying these isolates as MSSA [8]. Chromogenic media is a selective and differential media, which allows visual identification of target pathogens like MRSA on agar plates. Chromogenic agar incorporates chromogens that can be hydrolyzed by enzymes possessed by target pathogens, to produce desired colors. In order to reduce false positive identification, the media also incorporates inhibitors and antibiotics (to inhibit growth of unwanted organisms), or other chromogens to visually differentiate from target pathogens. Chromogenic agar for MRSA contains cefoxitin to eliminate the growth of MSSA, which allows for MRSA identification through positive growth on the medium. It is particularly useful as a screening tool, as it can be inoculated directly from swabs, or after an enrichment broth incubation. This has the ability to detect MRSA between 18 to 48 hours from acquiring clinical or screening samples [9, 10]. Laboratories typically use a combination of these techniques, as a way to confirm the results.

4.2.2 FTIR spectroscopy for microbial identification and discrimination based on antibiotic-resistance

FTIR spectroscopy is a rapid, and reagent-free microbial identification method. Like the MALDI-TOF MS, the technique is based on interrogating a spectral database to determine the identity of unknown isolates. FTIR spectra is the resultant fingerprint based on the biochemical makeup of the isolates. Various spectral regions can be exploited to analyze for species and sub-species level identification as well as strain typing [11-13]. Despite successful reports of microorganism identification using FTIR spectroscopy, there is a lack of report on successful discrimination between antibiotic resistant and sensitive microorganisms. Discrimination between glycopeptide intermediate and methicillin resistant *S. aureus* using FTIR spectroscopy combined with multivariate statistical analyses is one example of successful spectral discrimination based on antimicrobial resistance characteristics [14]. The use of machine-learning algorithms, support vector machine and artificial neural network were reported for discrimination between extended-

spectrum β-lactamase (ESBL) producing *Escherichia coli* and ESBL-negative isolates, as well as between *E. coli* isolates that are resistant and sensitive to antibiotics [15-17]. The limited number of studies published to date indicate the difficulty faced by FTIR spectroscopy in reliably discriminating antibiotic-resistant isolates from sensitives ones. Reported studies were lacking in sample size, and therefore is likely that the isolates used in the study did not encompass and represent the diversity in resistance-mechanisms that the isolates can possess, and the variability that pertain to non-resistance related sub-species characteristics [17, 18]. This study focused on whether identification of MRSA was possible through the combined use of a selective media and FTIR spectroscopy. The idea was to inhibit the growth of MSSA through the addition of cefoxitin into the media, and identifying *S. aureus*, from CoNS using the previously developed transflection FTIR spectral database. The results obtained from FTIR spectroscopy were compared to that of MALDI-TOF MS, the reference identification method, and false-positive and false-negative rates were determined for MRSA identification based on AST results based on minimum inhibition concentrations provided by VITEK 2 AST.

4.3 Materials and Methods

4.3.1 Clinical *Staphylococcus* isolates used for evaluation of the FTIR spectroscopy based identification of MRSA

Two hundred thirty five staphylococcal isolates (Table 4.1) from nasal screening, positive blood and sterile body fluid samples were prepared for FTIR spectral acquisition through two consecutive culturing from frozen glycerol stocks on blood agar plates (BAP). Identification results were provided by MALDI-TOF MS, and resistance to oxacillin were reported using VITEK 2 AST for staphylococcal isolates (bioMérieux SA, Marcy l'Étoile, France), to classify samples as methicillin-resistant or sensitive.

Table 4.1 Source and methicillin-resistance status of staphylococcal isolates used in the study

Genus	**Species**	**Resistance**	**Total Isolates**	**Positive Blood**	**Sterile Body Fluids**	**Nasal Screening**
Staphylococcus	*aureus*	MRSA	51	13	13	25
Staphylococcus	*aureus*	MSSA	32	16	16	0
Staphylococcus	*capitis*	MRS	6	5	1	0
Staphylococcus	*capitis*	MSS	19	11	8	0
Staphylococcus	*caprae*	MSS	2	1	1	0
Staphylococcus	*cohnii*	MRS	1	0	1	0
Staphylococcus	*epidermidis*	MRS	37	28	9	0
Staphylococcus	*epidermidis*	MSS	23	13	10	0
Staphylococcus	*haemolyticus*	MRS	10	4	6	0
Staphylococcus	*haemolyticus*	MSS	3	1	2	0
Staphylococcus	*hominis*	MRS	12	7	5	0
Staphylococcus	*hominis*	MSS	15	10	5	0
Staphylococcus	*lugdunensis*	MRS	3	1	2	0
Staphylococcus	*lugdunensis*	MSS	7	4	3	0
Staphylococcus	*simulans*	penicillin resistant	1	0	1	0
Staphylococcus	*species*	MRS	3	3	0	0
Staphylococcus	*species*	MSS	5	5	0	0
Staphylococcus	*warneri*	MSS	5	0	5	0
Total			**235**	**122**	**88**	**25**

4.3.2 Composition of media used for evaluation on the effect on FTIR spectroscopy based identification

Modified Columbia agar with 5 % sheep blood (BAP) was prepared as described in Table 4.2 using the base Columbia agar powder (Oxoid, Nepean, ON), defibrinated sheep blood (QuadFive, Ryegate, MT, USA), cefoxitin salt (MilliporeSigma, St. Louis, MN, USA), and Strep/Staph Supplement (Oxoid, Nepean, ON). Additionally, commercially prepared BAP (Oxoid, Nepean, ON), and BBL ChromAgar MRSA II (Becton, Dickenson and Company, Franklin Lakes, New Jersey, USA) were used, to compare the spectral features and confirm antibiotic resistance respectively.

Table 4.2 Formula of tested culture media in the study

Code	**Base**	**[FOX] (µg/ml)**	**[CNA]**
BAPO	Columbia with 5 % Sheep blood	0	0
BAPT	Columbia with 5 % Sheep blood	0	0
4FOX	Columbia with 5 % Sheep blood	4	0
4FOX_CNA	Columbia with 5 % Sheep blood	4	15 µg/ml Nalidixic Acid 10 µg/ml Colistin
8FOX	Columbia with 5 % Sheep blood	8	0
8FOX_CNA	Columbia with 5 % Sheep blood	8	15 µg/ml Nalidixic Acid 10 µg/ml Colistin
CNAO	Columbia with 5 % Sheep blood	0	10 µg/ml Nalidixic Acid 10 µg/ml Colistin
CNAT	Columbia with 5 % Sheep blood	0	15 µg/ml Nalidixic Acid 10 µg/ml Colistin
Chrome	BD BBL ChromAgar MRSA II	5.2	7.52mg combined

4.3.3 FTIR spectral acquisition and processing parameters used for microbial identification

Using sterile 1µl disposable loops, isolated colonies were deposited onto IR reflective E-glass (Kevley Technologies, Chesterland, OH, USA) as thin films and air-dried. Using the Cary 630 FTIR spectrometer (Agilent Technologies, Santa Clara, CA, USA) coupled with a SurveyIR microscope (Czitek, Danbury, CT, USA) with 1 mm aperture, triplicate spectra were acquired per isolate, per media type. Using MicroLab Software (Agilent Technologies, Santa Clara, CA, USA), Triplicate spectra were acquired per isolate with 64 coadded scans at 8 cm^{-1} resolution and zero-filling with a factor of 8, in the spectral range 4000-650 cm^{-1} per spectrum. A background spectrum was acquired from a bare space on E-glass with the same spectral parameters, between spectral acquisition. FTIR spectra were processed using first derivative and vector normalization to remove effects from baseline shifts and sample thickness variability respectively. Genus and species level identification was achieved through the use of a previously developed transflection FTIR spectral database.

4.3.4 FTIR spectral analysis: the effect of the presence of antibiotics in culture media on the spectral features of staphylococcal isolates, and identification at the genus and species level.

The effect of media composition modification on the FTIR spectral fingerprints of the staphylococcal isolates was determined using principal component analysis (PCA), hierarchical cluster analysis (HCA) and forward search algorithm. Forward search algorithm identified spectral regions that optimized separation between groups of interest (i.e. FTIR spectra of isolates cultured with and without presence of antibiotics). The extent of spectral discrimination between isolates cultured with and without antibiotics were visualized using HCA and PCA, in dendrograms and PC plots respectively, using broad and specific spectral regions that were identified by the forward search algorithm. Isolates were identified as *S. aureus* or CoNS, using a previously developed transflection-FTIR spectral database in the previous chapter, based on spectral similarity to those in the database. Isolates with discrepant species identity between MALDI-TOF MS and FTIR spectroscopy were reidentified by both techniques for confirmation.

4.4 Results and Discussion

4.4.1 Effect of cefoxitin and CNA on FTIR spectra of microorganisms

The FTIR spectral database developed previously for species identification was based on isolates cultured on BAP that were commercially prepared by Oxoid Canada (Thermo Fisher Scientific, ON, Canada). One of the concerns with the media modification with the addition of antibiotics was that there would be substantial changes reflected in the bacteria's FTIR spectral profile, as a result of changed metabolism, that may hinder accurate genus and species level identification. The identification capability of the spectral database with isolates that were cultured on modified BAP has not been done prior to this study. The FTIR spectra acquired from isolates cultured on media with and without antibiotics were all identified using the previously developed transflection-FTIR spectral database. Based on prior identification results for *S. aureus* isolates (*n* = 891, 98.7 %), FTIR spectral identification of isolates cultured on Oxoid BAP (BAPO) were considered reliable, and used for comparison in identification performance to those cultured on the modified media.

Using PCA and HCA, FTIR spectra acquired from BAPO and BAPT were compared in supervised and unsupervised analyses (Figure 4.1). Based on the broad spectral region 1350-800 cm^{-1}, the PC score plot of PC 1 vs PC 2 showed that there were no clear spectral clustering between *S. aureus* cultured on the two differently prepared BAP. Although spectral regions for optimized spectral discrimination was identified with the forward search algorithm, there were no clear spectral distinction that could be made between the two groups (Figure 4.2). The manual BAP media preparation was not a contributing variable in the differences observed between spectra acquired from colonies cultured with and without the presence of antibiotics in the following analyses.

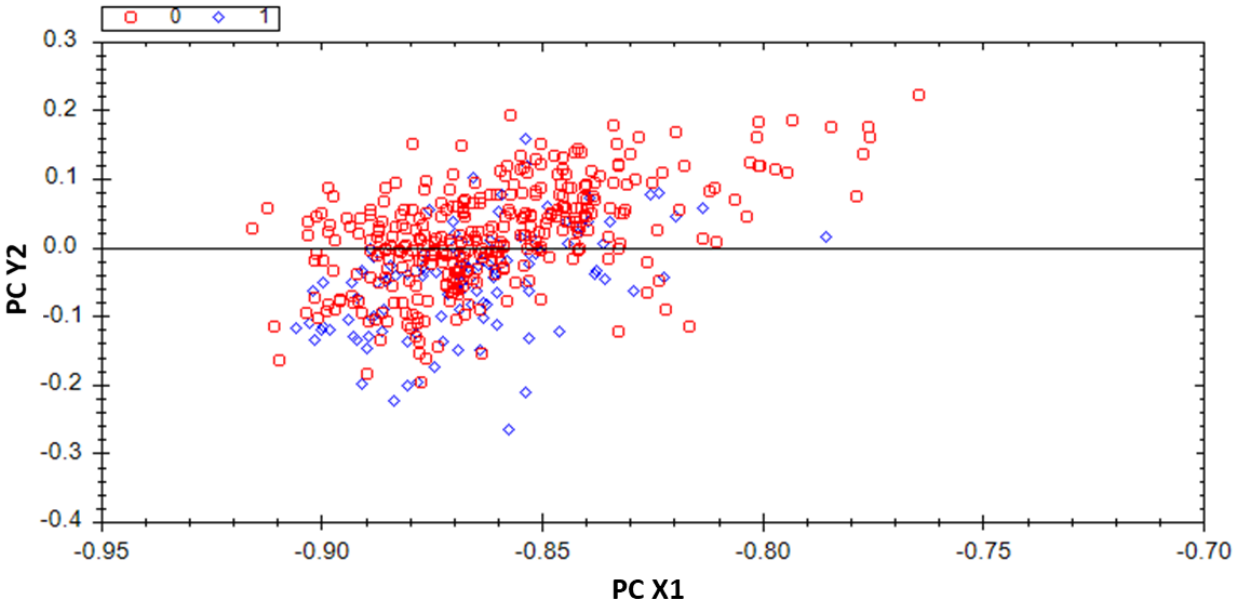
The effect of cefoxitin presence on the MRSA spectral profile was determined in a pairwise comparison of spectra that were acquired from BAP with and without cefoxitin. As per Figure 4.3, the difference between FTIR spectra of MRSA cultured on BAP with and without cefoxitin (4 µg/ml) were not enough to distinguish the two groups in PCA using broad spectral region. No distinct spectral clustering were observed when using identified spectral regions from forward search algorithm. This was also confirmed in HCA, where the generated dendrogram were constructed using cosine distance and Ward linkage. Schelli et al. reported metabolic changes for both MSSA and MRSA when comparing cultivation with and without sub-lethal doses of

methicillin, indicating that the presence of methicillin during growth affected several metabolic pathways in *S. aureus*. The differences observed in MSSA isolates' metabolism was much more dramatic compared to that of MRSA isolates, speculating that the MRSA were likely less susceptible to metabolic changes due to the possession of resistance mechanisms (*SCCmec* and PBP2a system) [19].

The same analyses were employed to determine whether the presence of CNA altered the spectral characteristics of *S. aureus* isolates. It was confirmed that there was a lack of differences between the FTIR spectral profiles between *S. aureus* isolates cultured on BAPO and BAP with CNA (CNAT) (Figure 4.4). Colistin is an antibiotic that binds to phosphate groups of lipid A on Gram-negative bacteria, through electrostatic interaction, and therefore, is not known to interact with Gram-positive bacteria. *Staphylococcus aureus* is intrinsically resistant to colistin, via the presence of genes that encode subunits of ATP synthase [20]. Studies have indicated that the presence of colistin in media during growth of *S. aureus* affects their characteristics, such as increase in autolytic activity, decrease in positive charge in the cell surface and decrease in ions such as Na^{+}, Mg^{2+}, K^{+} (increased leakage of ions) [21]. Nalidixic acid is a broad-spectrum, quinolone antibiotic, that is used against Gram-negative bacteria. A mutation in the *grlA* of *S. aureus*, which encodes topoisomerase IV (A_2B_2 enzyme), is known to implicate nalidixic acid resistance through decreased levels of binding to the enzyme [22, 23]. Overexpression of efflux pumps (such as NorA) that prevent introduction of nalidixic acid and other quinolones into the bacteria may occur [23]. However, PCA and HCA indicate that the changes caused by the presence of colistin and nalidixic acid did not affect the FTIR spectra to distinguish those from FTIR spectra acquired from isolates cultured without their presence.

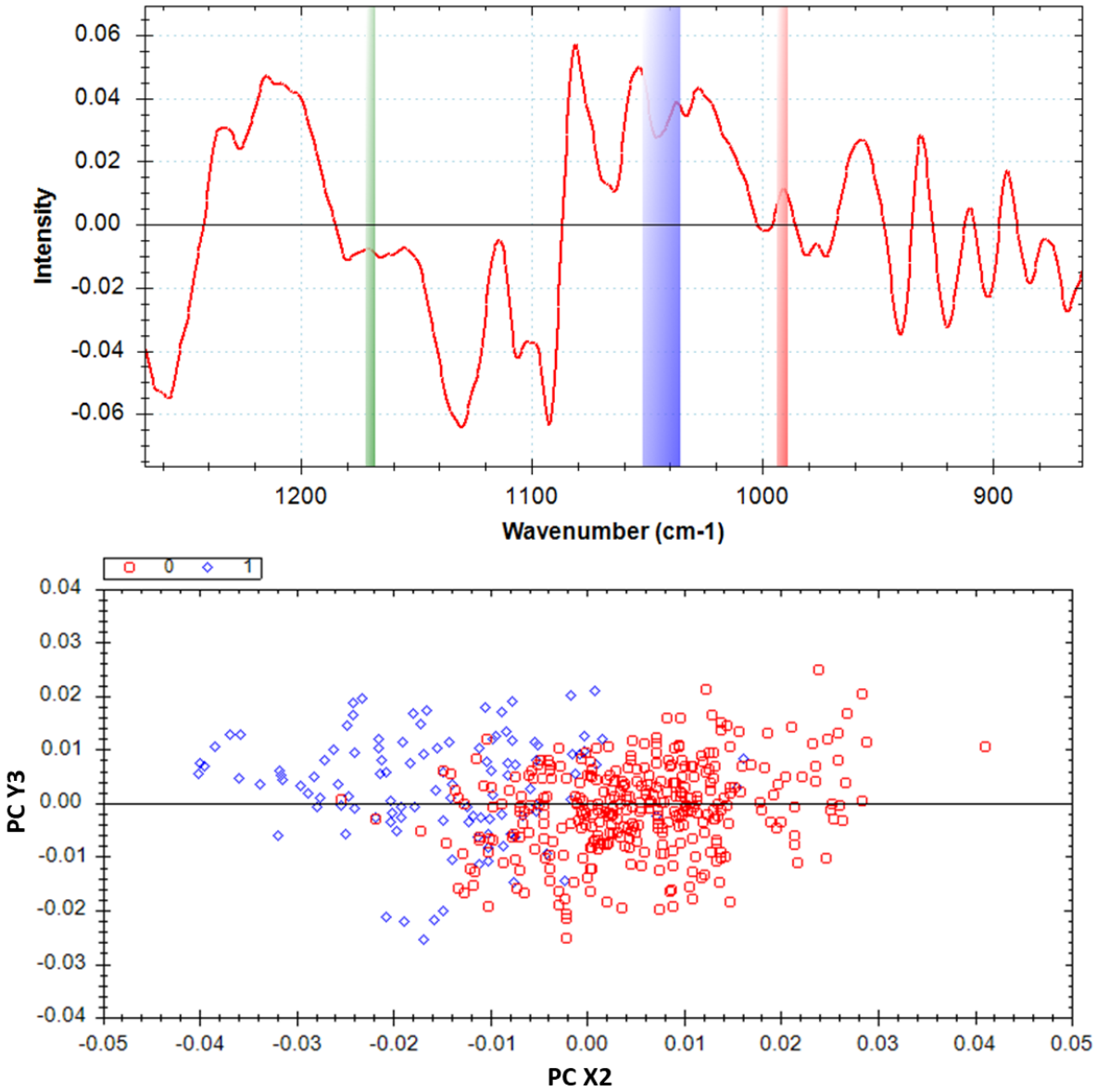
Despite the metabolic and structural differences that may result from presence of these antibiotics, their addition to BAP did not affect the identification of the *Staphylococcus* isolates based on the use of transflection-FTIR spectral database, at the level of identifying CoNS and *S. aureus* (Table 4.3).

Figure 4.1 PCA in spectral region 1350-800 cm^{-1}, showcasing a lack of spectral differences between FTIR spectra of S. aureus isolates cultured on Oxoid manufactured (Red circle) and self-prepared (Blue diamond) BAP in score plot for PC 1 vs PC 2



Score plot PC1 vs PC2 shows that transflection FTIR spectra acquired from *S. aureus* isolates cultured on pre-prepared Oxoid BAP and self-prepared BAP do not have differences in their spectra when analyzed unsupervised in spectral region 1350-800 cm^{-1}. This indicated that the BAPT were comparable to the commercially prepared counterpart, passing the quality check in terms of properly preparing the agar plates.

Figure 4.2 PCA in selected spectral region (989-994,1036-1052,1168-1172 cm^{-1}, highlighted in top image), showcasing the lack of spectral differences between FTIR spectra of S. aureus *isolates cultured on Oxoid manufactured (red circle) and self-prepared (blue diamond) BAP in score plot for PC 2 vs PC 3 (bottom).*



Spectral differences in the selected regions (highlighted in the spectrum in top figure) were not clear, although some separation can be observed in the score plot (PC2 vs PC3). This further confirms that the media preparation method did not affect the FTIR spectral characteristics, and thus is not a variable that need to be considered when comparing FTIR spectra of isolate cultured on media with and without antibiotics.

Figure 4.3 PCA in broad (1350-800 cm^{-1}, top figure) and selected spectral regions (926-936,966-972,1030-1036,1058-1070,1088-1096,1174-1186,1276-1292 cm^{-1}, bottom figure, highlighted in middle figure)showcasing the lack of spectral differences between FTIR spectra acquired from MRSA isolates cultured on BAP with (blue diamond) and without (red circle) 4 µg/ml cefoxitin in score plot for PC 2 vs PC 3.

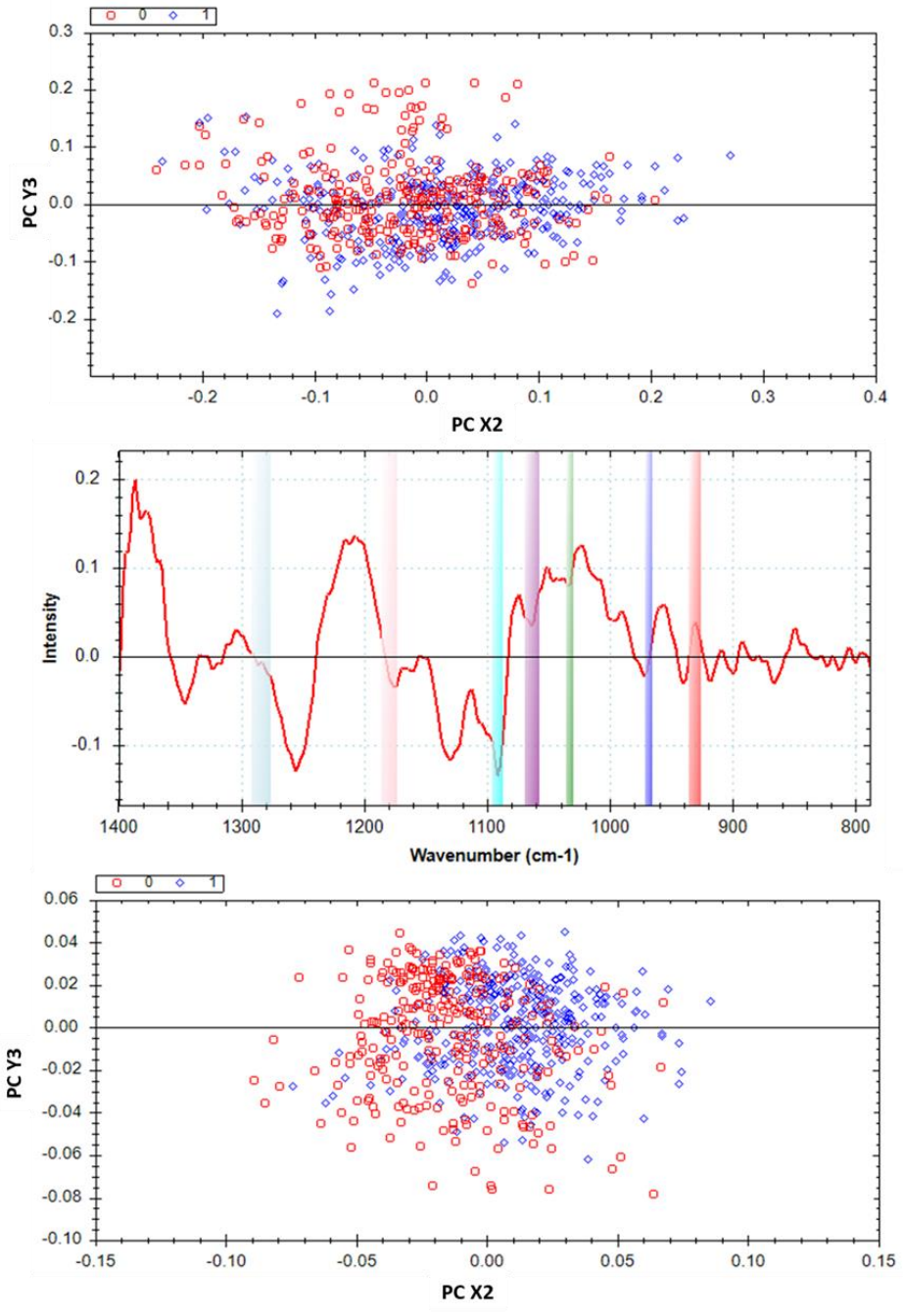
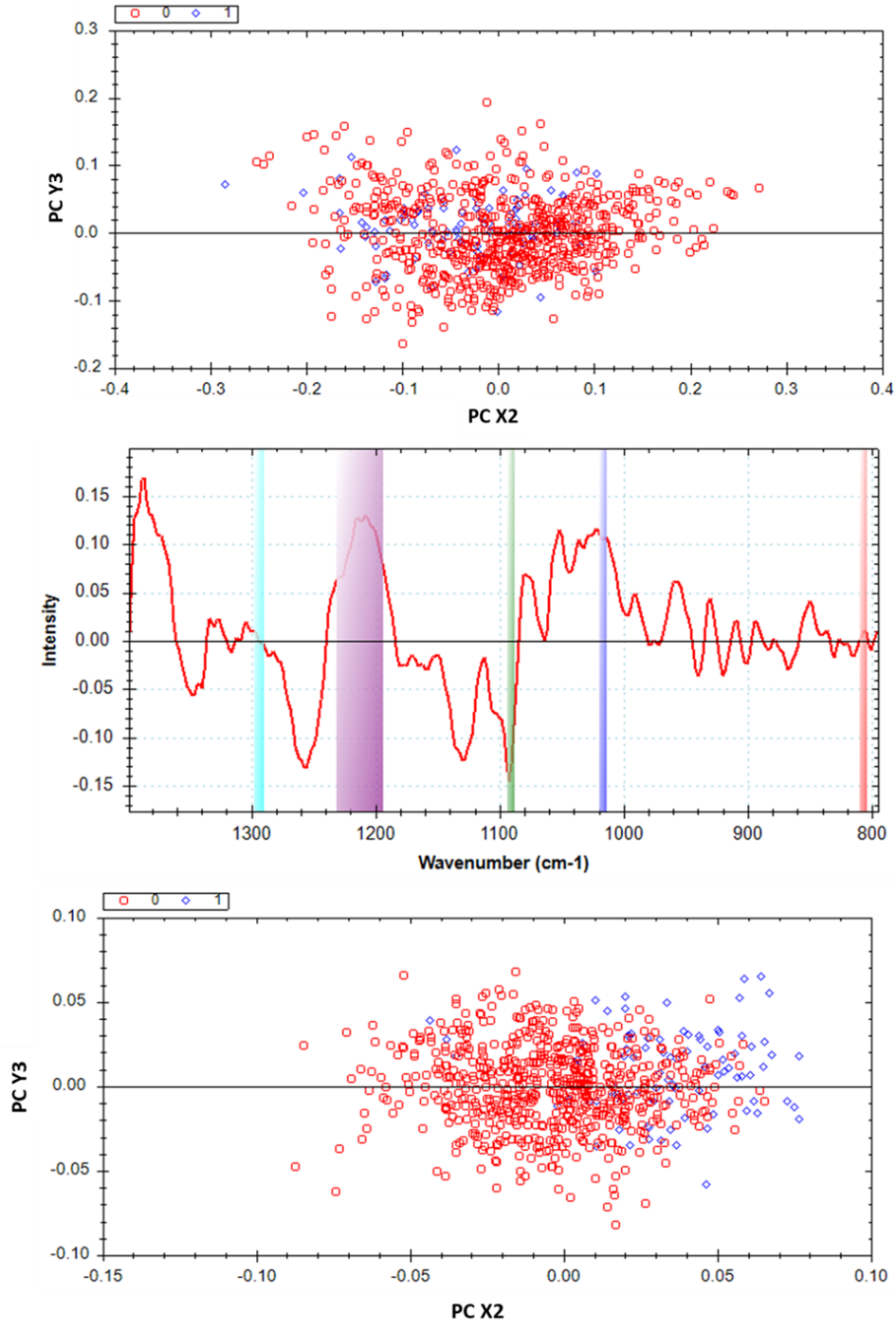


Figure 4.4 PCA in broad (1350-800 cm^{-1}, top figure) and selected (804-810,1014-1020,1088-1094,1194-1232,1290-1298 cm^{-1} bottom figure, highlighted in middle figure) spectral region, showcasing the lack of spectral discrimination between FTIR spectra acquired from MRSA isolates cultured on BAP with (blue diamond) and without (red circle) CNA in score plot for PC 2 vs PC 3.



4.4.2 Identification of staphylococcal species using FTIR spectroscopy

A total of 4200 spectra were collected from 229 staphylococcal isolates, cultured on three different media types, BAP without any antibiotics (control), with 4 µg/ml cefoxitin and 15 µg/ml CNA, and with 8 µg/ml cefoxitin and 15 µg/ml CNA. The identity of the isolates that had positive growth on the media were predicted using a transflection-FTIR spectral database. One isolate was identified as *Escherichia coli* and another was predicted as *Enterococcus faecalis*. These isolates were re-identified by both FTIR spectroscopy and MALDI-TOF MS, and was confirmed that the original FTIR spectral identification results were correct. These isolates were therefore removed from subsequent analyses and statistics for sensitivity and specificity for MRSA detection. Overall, 100 % correct identification of *S. aureus* was achieved on all media (data only shown in table for three media types; Table 4.3). Furthermore, 100 %, 98.1 % and 100 % of the CoNS were correctly identified as CoNS when cultured on BAP, CNA_4FOX and CNA_8FOX respectively. All *S. aureus* isolates on all media were correctly identified as such, resulting in 100 % correct identification, in reference to MALDI-TOF MS. At the species level for CoNS species, BAP, CNA_4FOX and CNA_8FOX media had 61.7 %, 28.8 % and 30 % concordance to MALDI-TOF MS results. Although MRSA identification using FTIR spectra acquired from media containing cefoxitin, colistin and nalidixic acid, the species level identification with CoNS performed lower for CNA_4FOX and CNA_8FOX, relative to BAP.

Table 4.3 Summary of identification results for S. aureus and CoNS using transflection FTIR spectral database, from isolates cultured on BAP with and without antibiotics

	BAP			**CNA_4FOX**			**CNA_8FOX**		
Species	*n*	Correct	(%)	*n*	Correct	(%)	*n*	Correct	(%)
S. aureus	99	99	100	58	58	100	51	51	100
CoNS	130	130	100	53	50	94.34	48	42	87.5
Total	229	229	100	111	108	97.3	99	93	93.9

100 % of *S. aureus* isolates were identified correctly on all three media types, whereas 100 %, 85.4% and 87.5 % of CoNS isolates were correctly identified on BAP, CNA_4FOX and CNA_8FOX respectively. Overall, the spectral database had 100 %, 97.3 % and 93.9 % correct identification using the transflection FTIR spectral database.

4.4.3 MRSA detection using modified agar in combination with FTIR spectroscopy

Sensitivity and specificity were 100 % and 95.3 % respectively for *S. aureus* (*n* = 99), using 4 µg/ml cefoxitin agar plates at 24 hours incubation. This was shown to be better than 8 µg/ml cefoxitin agar plates, which had 100 % specificity, but 92.9 % sensitivity (*n* = 52), and 7.1 % (*n* = 5) very major error. Very major error is defines as false negative result, where MRSA is misidentified as MSSA. Major error is the false positive error, where MSSA isolates are identified as MRSA. Clinically, a very major error implicates that ineffective treatment course may be implemented to a patient, which may worsen their health condition due to the lack of appropriate antibiotic therapy choice. A major error in the clinical world implies that a strong antibiotic that perhaps may be considered a last-line of defense is used unnecessarily. The excessive use of antibiotics is one of the causes of increased prevalence of antibiotic resistant microorganisms. Therefore improvement in appropriate use of antibiotics, as well as reduction in unnecessary use are important part of preventing further increase in antibiotic resistant microorganisms [4, 24] . Chromogenic agar had 92.9 % sensitivity, resulting in 7.1 % very major error (*n* = 4), and 100 % specificity. After 48 hours of incubation, the sensitivity of MRSA detection improved to 100 %, 94.6 % and 100 % for CNA_4FOX, CNA_8FOX and chrome agar respectively. After 48 hours of incubation all MRSA isolates had positive growth on CNA_4FOX and chrome agar. Of the 4 MRSA isolates that did not grow on CNA_8FOX after 24 hours of incubation, three isolates were *mecC* MRSA, and these were the only three that did not show growth after 48 hours of incubation.

It is important to note that these isolates had growth on CNA_4FOX media at both 24 and 48 hours incubation. *mecC* MRSA tend to have lower resistance level against oxacillin and cefoxitin, and therefore the concentration of cefoxitin present in the media at 8 µg/ml may have been sufficient to prevent the growth of *mecC* MRSA isolates [25]. Additionally, two of the three *mecC* MRSA isolates that failed to grow on CNA_8FOX lacked growth after 24 hours of incubation on chrome agar. The specificity of the CNA_8FOX remained at 100 %, but the specificity of CNA_4FOX and chrome agar decreased from 95.3 % to 88.4 % and 100 % to 97.7 % respectively. The MSSA isolate that had positive growth on chrome agar after 48 hours incubation showed to be white colonies, and thus not MRSA, but likely a MR-CoNS isolate. MSSA isolates that had positive growth on the CNA_4FOX media after 24 and 48 hours incubation were all identified as *S. aureus*. Overall, CNA_4FOX agar plates had the best performance for correctly detecting MRSA isolates, with a categorical agreement of 98 % after 24 hours of incubation. Chrome agar performed the best overall (categorical agreement) after 48 hours, with 99 %. However, the modified media both did well, with 98 % and 949 % categorical agreement as well. Based on the ISO accepted very major error and major error rates (≤1.5 % and ≤3 % respectively) for a new method [24], CNA_4FOX obtained acceptable results, in sensitivity, with very major error rates less than 1.5 % at both 24 and 48 hours. However, it was not able to meet the standard for major error rates, resulting in 4.7 % major error rates. CNA_8FOX on the other hand met the standard for major error, but not the very major error rates (8.9 %). Based on these results, the CNA_4FOX media would provide the most reliable presumptive MRSA results, prior to obtaining MIC results by other methods.

Compared to the high sensitivity and specificity for MRSA detection, both the sensitivity and specificity for MR-CoNS were lower after both 24 and 48 hours incubation. The results were not surprising, as CLSI recommends the detection of methicillin resistance in CoNS except *S. lugdunensis* by using disk diffusion test using oxacillin or cefoxitin, or MIC from oxacillin [26]. Cefoxitin MIC is regarded to be inferior compared to disk diffusion using 30 µg cefoxitin by EUCAST [27], for all CoNS species with the exception of *S. lugdunensis*. Despite the use of cefoxitin MIC as an accepted method for determining methicillin resistance in *S. lugdunensis*, in this study, out of three MRS-*S. lugudensis*, only one isolate had positive growth on CNA_4FOX after 24 hours incubation, whereas, the other two isolates did not grow on any media even after 48 hours of incubation. CNA_4FOX had the best overall performance, with 91.4 % sensitivity and

100 % specificity after 24 hours incubation for MRS-CONS. CNA-8FOX and chrome agar performed worse, at 85.5 % and 60 % sensitivity respectively, and 100 % specificity for both media. Since BD's BBL ChromAgar MRSA II is a selective media for MRSA screening, as expected, there was low sensitivity and high very major error rates, at both 24 and 48 hours for detecting MR-CoNS (Table 4.4 and Table 4.5). This is due to the presence of inhibitory agents in the media composition, to favor the isolation of MRSA from screening samples. After 48 hours of incubation, one MS-*S. epidermidis* showed positive growth, but was identified as MRSA, rather than MR-*S. epidermidis.* Overall the specificity after 48 hours for MR-CONS identification was 87.7 %, 100 % and 98.6 % for CNA_4FOX, CNA_8FOX and chrome agar respectively. The overall categorical agreement for MR-CONS on CNA_4FOX, CNA_8FOX and chrome agar were 91.4 %, 93.8 %, and 82.9 % respectively.

Table 4.4 Positive growth at 24 and 48hrs of incubation on cefoxitin containing media for staphylococcal isolates

		CNA_4FOX					CNA_8FOX					Chrome				
Species		*n*	24 hrs	48 hrs	(%) 24 hrs	(%) 48 hrs	*n*	24 hrs	48 hrs	(%) 24 hrs	(%) 48 hrs	*n*	24 hrs	48 hrs	(%) 24 hrs	(%) 48 hrs
S. aureus	MRSA	56	56	56	100	100	56	51	53	91.1	94.6	56	52	56	92.9	100
S. capitis	MRS	5	5	5	100	100	5	5	5	100	100	5	3	3	60	60
S. cohnii	MRS	1	1	1	100	100	1	1	1	100	100	1	1	1	100	100
S. epidermidis	MRS	27	27	27	100	100	27	26	27	96.3	100	27	15	18	55.6	66.7
S. haemolyticus	MRS	9	9	9	100	100	9	8	9	88.9	100	9	8	8	88.9	88.9
S. hominis	MRS	10	7	8	70	80	10	7	7	70	70	10	6	6	60	60
S. lugdunensis	MRS	3	1	1	33.3	33.3	3	0	0	0	0	3	0	0	0	0
S. species	MRS	1	1	1	100	100	1	1	1	100	100	1	0	0	0	0
S. aureus	MSSA	43	2	2	4.7	4.7	43	0	0	0	0	43	0	1	0	2.3
S. capitis	MSS	19	0	1	0	5.3	19	0	0	0	0	19	1	1	5.3	5.3
S. caprae	MSS	2	0	0	0	0	2	0	0	0	0	2	0	0	0	0
S. epidermidis	MSS	22	0	0	0	0	22	0	0	0	0	22	0	0	0	0
S. haemolyticus	MSS	3	0	0	0	0	3	0	0	0	0	3	0	0	0	0
S. hominis	MSS	15	0	0	0	0	15	0	0	0	0	15	0	0	0	0
S. lugdunensis	MSS	7	0	0	0	0	7	0	0	0	0	7	0	0	0	0
S. species	MSS	2	0	0	0	0	2	0	0	0	0	2	0	0	0	0
S. warneri	MSS	4	0	0	0	0	4	0	0	0	0	4	0	0	0	0
True Positive Rate		112	107	108	95.5	96.4	112	99	103	88.4	92	112	85	92	75.9	82.1
False Negative Rate		112	5	4	4.5	3.6	112	13	9	11.6	8	112	27	20	24.1	17.9
True Negative Rate		117	115	112	98.3	95.7	117	117	117	100	100	117	116	114	99.1	97.4
False Positive Rate		117	2	3	1.7	2.6	117	0	0	0	0	117	1	2	0.9	1.7
Categorical Agreement		229	222	220	96.9	96.1	229	216	220	94.3	96.1	229	201	206	87.8	90

Table 4.5 Sensitivity, specificity, and categorical agreement for S. aureus *and CoNS cultured on antibiotic containing media, in comparison to chromogenic agar*

	CNA_4FOX				CNA_8FOX				Chrome			
S. aureus	*n* (24 hrs)	%	*n* (48 hrs)	%	*n* (24 hrs)	%	*n* (48 hrs)	%	*n* (24 hrs)	%	*n* (48 hrs)	%
Sensitivity	56	100.0	56	100.0	51	91.1	53	94.6	52	92.9	56	100.0
Specificity	41	95.3	41	95.3	43	100.0	43	100.0	43	100.0	42	97.7
Very Major Error	0	0.0	0	0.0	5	8.9	3	5.4	4	7.1	0	0.0
Major Error	2	4.7	2	4.7	0	0.0	0	0.0	0	0.0	1	2.3
Categorical Agreement	97	98.0	97	98.0	94	94.9	96	97.0	95	96.0	98	99.0
	CNA_4FOX				**CNA_8FOX**				**Chrome**			
CoNS	*n* (24 hrs)	%	*n* (48 hrs)	%	*n* (24 hrs)	%	*n* (48 hrs)	%	*n* (24 hrs)	%	*n* (48 hrs)	%
Sensitivity	50	90.9	51	92.7	47	85.5	49	89.1	33	60.0	36	65.5
Specificity	74	100.0	73	98.6	74	100.0	74	100.0	74	100.0	73	98.6
Very Major Error	5	9.1	4	7.3	8	14.5	6	10.9	22	40.0	19	34.5
Major Error	0	0.0	1	1.4	0	0.0	0	0.0	0	0.0	1	1.4
Categorical Agreement	124	96.1	124	96.1	121	93.8	123	95.3	107	82.9	109	84.5

4.5 Conclusion

A bi-plate identification concept with FTIR spectroscopy where staphylococcal isolates were cultured on BAP and BAP with cefoxitin and CNA was tested for its MRSA identification capabilities. The selective growth on cefoxitin containing media allowed identification of methicillin resistant isolates, while the growth of methicillin sensitive isolates were inhibited. The transflection FTIR spectral database correctly identified 100 % (*n* = 99) *S. aureus* from CoNS (*n* = 130) on media both with and without cefoxitin, enabling the identification of MRSA. In this study, 98 % categorical agreement, with 0 % very major error and 4.7 % major error was achieved after 24 hours incubation, using BAP containing 4 µg/ml cefoxitin for MRSA identification. This media variation also had the best categorical agreement for MRS-CoNS as well at 96.1 %. Although it is not suitable for MRS-CoNS identification, the method provides as a reliable screening step in identifying presumptive MRSA isolates, which can later be confirmed through molecular and phenotypic methods. In the long-term this may also contribute to antimicrobial stewardship, as it helps reduce overuse of antibiotics through faster determination of the methicillin resistance of staphylococcal isolates. This in turn can contribute to slowing down emergence of microorganisms that are more resistant to one or more antibiotics. The presumptive identification of MRSA should be confirmed with AST results the following day, along with the AST profiles for other classes of antibiotics. Furthermore, the use of this modified agar in combination with FTIR spectroscopy, provides a cheaper alternative to using chromogenic agar plates. An extended study with more strain types and source variation should be conducted to evaluate whether there are other limiting factors that arise that were not observed in the current study. The advantage of the method is that isolate characterization at species and subspecies level can be carried out on the same acquired FTIR spectral data. This is particularly beneficial for infection control as it shortens the time required until isolates are identified for further strain typing analyses. This modified agar approach should be evaluated for other pathogens of interest such as vancomycin resistant enterococci and carbapenem resistant *Enterobacteriaceae*, by adding the antibiotics of interest in the generic blood agar media.

4.6 References

1. Kourtis, A.P., et al., *Vital Signs: Epidemiology and Recent Trends in Methicillin-Resistant and in Methicillin-Susceptible Staphylococcus aureus Bloodstream Infections—United States.* Morbidity and Mortality Weekly Report, 2019. **68**(9): p. 214.
2. Organization, W.H., *Antimicrobial resistance: global report on surveillance*. 2014: World Health Organization.
3. Control, C.f.D. and Prevention, *Antibiotic resistance threats in the United States, 2013*. 2013: Centres for Disease Control and Prevention, US Department of Health and ….
4. CDC, *Antibiotic resistance threats in the United States, 2019*. 2019, U.S. Department of Health and Human Services: Atlanta, GA.
5. Lévesque, S., et al., *A side by side comparison of Bruker Biotyper and VITEK MS: utility of MALDI-TOF MS technology for microorganism identification in a public health reference laboratory.* PloS one, 2015. **10**(12): p. e0144878.
6. Zhang, K., et al., *New quadriplex PCR assay for detection of methicillin and mupirocin resistance and simultaneous discrimination of Staphylococcus aureus from coagulase-negative staphylococci.* Journal of clinical microbiology, 2004. **42**(11): p. 4947-4955.
7. Bard, J.D. and F. Lee, *Why can't we just use PCR? The role of genotypic versus phenotypic testing for antimicrobial resistance testing.* Clinical microbiology newsletter, 2018. **40**(11): p. 87-95.
8. Ford, B.A., *mecC-harboring methicillin-resistant Staphylococcus aureus: hiding in plain sight.* Journal of clinical microbiology, 2018. **56**(1).
9. Perry, J.D., *A decade of development of chromogenic culture media for clinical microbiology in an era of molecular diagnostics.* Clinical microbiology reviews, 2017. **30**(2): p. 449-479.
10. Nonhoff, C., et al., *Comparison of three chromogenic media and enrichment broth media for the detection of methicillin-resistant Staphylococcus aureus from mucocutaneous screening specimens.* European journal of clinical microbiology & infectious diseases, 2009. **28**(4): p. 363-369.
11. Amiali, N.M., et al., *Epidemiological typing of methicillin-resistant Staphylococcus aureus strains by Fourier transform infrared spectroscopy.* Journal of microbiological methods, 2007. **69**(1): p. 146-153.
12. Lam, L.M., et al., *Reagent-Free Identification of Clinical Yeasts Using Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy.* Journal of clinical microbiology, 2019: p. JCM. 01739-18.
13. Amiali, N.M., et al., *Rapid identification of coagulase-negative staphylococci by Fourier transform infrared spectroscopy.* Journal of microbiological methods, 2007. **68**(2): p. 236-242.
14. Amiali, N.M., et al., *Evaluation of Fourier transform infrared spectroscopy for the rapid identification of glycopeptide-intermediate Staphylococcus aureus.* Journal of antimicrobial chemotherapy, 2007. **61**(1): p. 95-102.
15. Sharaha, U., et al., *Detection of Extended-Spectrum β-Lactamase-Producing Escherichia coli Using Infrared Microscopy and Machine-Learning Algorithms.* Analytical chemistry, 2019. **91**(3): p. 2525-2530.

16. Salman, A., et al., *Detection of antibiotic resistant Escherichia Coli bacteria using infrared microscopy and advanced multivariate analysis.* Analyst, 2017. **142**(12): p. 2136-2144.
17. Lechowicz, L., et al., *The use of infrared spectroscopy and artificial neural networks for detection of uropathogenic Escherichia coli strains' susceptibility to cephalothin.* Acta Biochimica Polonica, 2013. **60**(4).
18. Novais, Â., et al., *Fourier transform infrared spectroscopy: unlocking fundamentals and prospects for bacterial strain typing.* European Journal of Clinical Microbiology & Infectious Diseases, 2019. **38**(3): p. 427-448.
19. Schelli, K., et al., *Staphylococcus aureus methicillin resistance detected by HPLC-MS/MS targeted metabolic profiling.* Journal of Chromatography B, 2017. **1047**: p. 124-130.
20. Vestergaard, M., et al., *Inhibition of the ATP synthase eliminates the intrinsic resistance of Staphylococcus aureus towards polymyxins.* MBio, 2017. **8**(5).
21. Si, W., et al., *Colistin induces S. aureus susceptibility to bacitracin.* Frontiers in microbiology, 2018. **9**: p. 2805.
22. Griggs, D.J., K. Gensberg, and L. Piddock, *Mutations in gyrA gene of quinolone-resistant Salmonella serotypes isolated from humans and animals.* Antimicrobial Agents and Chemotherapy, 1996. **40**(4): p. 1009-1013.
23. Ruiz, J., *Mechanisms of resistance to quinolones: target alterations, decreased accumulation and DNA gyrase protection.* Journal of Antimicrobial Chemotherapy, 2003. **51**(5): p. 1109-1117.
24. Reller, L.B., et al., *Antimicrobial susceptibility testing: a review of general principles and contemporary practices.* Clinical infectious diseases, 2009. **49**(11): p. 1749-1755.
25. Paterson, G.K., E.M. Harrison, and M.A. Holmes, *The emergence of mecC methicillin-resistant Staphylococcus aureus.* Trends in microbiology, 2014. **22**(1): p. 42-47.
26. CLSI, *Performance Standards for Antimicrobial Disk Susceptibility Tests; Approved Standard - Twelfth Edition*, C.a.L.S. Institute, Editor. 2015.
27. Testing, E.C.o.A.S., *The European Committee on Antimicrobial Susceptibility Testing. Breakpoint tables for interpretation of MICs and zone diameters.*

Version 11.0, 2021. 2021.

Connecting Statement

Successful identification of MRSA was achieved through the combined use of antibiotic containing media and FTIR spectral database. To further evaluate the discriminatory capabilities of transflection-FTIR spectroscopy, in the next chapter, the technique was subjected to sub-species level discrimination, focusing on discrimination between strain-types, using supervised and unsupervised multivariate statistical techniques using VRE isolates collected from patient screening samples.

Chapter 5. FTIR Spectroscopy as Alternative Strain Typing Technique for Infection Control.

5.1 Abstract

Whole-organism fingerprinting by transflection-FTIR spectroscopy is a rapid, reagent-free technique for bacterial identification and classification with subspecies-level discriminatory capabilities. Rapid strain typing is useful in surveying and detecting nosocomial outbreaks, and can help infection control team, to implement necessary protocols to intercept the spread of pathogens to at-risk patients. Pulsed-field gel electrophoresis (PFGE) is the current gold-standard method for strain typing in outbreak investigations for many organisms including vancomycin resistant *Enterococcus faecium* (VRE). Transflection-FTIR spectroscopy was evaluated for its VRE subspecies level discrimination capabilities in reference to PFGE clustering results, in retrospective (n = 214) and prospective (n = 129) studies that analyzed samples collected from environmental and patient screenings over a 18 and seven month period respectively. Spectral discrimination between the two most common pulsotypes, AA (n = 34) and CC (n = 109) was demonstrated by principal component analysis (PCA) in combination with support vector machine and unsupervised PCA-linear discriminant analysis (PCA-LDA). Discrimination resulted in four (2.8 %) outliers. However, upon re-analysis of these samples by PFGE, the FTIR results for three of these outliers were found to be correct, improving the concordance between the two methods to 99.3 %. A cut-off value for determining spectrally "clonal" isolates was determined using squared Mahalanobis distances calculated between each isolate in PCA-LDA. Based on the developed method, AA and CC isolates respectively had 91.7 % and 91 % categorical agreement with PFGE. In both retrospective and prospective studies, the FTIR spectrotyping method had high true positive rates relative to PFGE pulsotype clustering, showcasing that isolates of the same pulsotype tend to also be spectrally indistinguishable. Discrepancies between PFGE and FTIR spectrotyping were also observed, likely because FTIR spectra represent the metabolic and structural status of the organism, while PFGE is based on macrofragments of the DNA. Overall, the method was able to identify clusters prospectively, with analyses performed per unit, with the addition of new samples on an incoming basis. At the end of the prospective study, 8 and 15 clusters were identified in 5 and 12 wards respectively. Clustering between environmental and patient isolates that originated from the same ward by spectrotyping demonstrated that the sample source does not affect the analysis, and has potential in tracing transmission routes. The ease of the method enables

more samples and data to be acquired to have a broader understanding on the VRE strain type evolution within patients, between patients and in the environment. Furthermore, when combined with a transflection-FTIR spectral database for species identification, both identification and strain typing results can be achieved from the same set of FTIR spectra acquired from samples, thereby reducing the time and cost required for outbreak investigation. The results in this study demonstrated that transflection-FTIR spectrotyping can be considered for routine use, with high throughput capabilities and improved turnaround time to results relative to PFGE.

5.2 Introduction

Infection prevention and control is critical in minimizing hospital acquired infections, especially those caused by antibiotic resistant organisms, such as vancomycin resistant *Enterococcus faecium* (VRE), methicillin resistant *Staphylococcus aureus* (MRSA) and carbapenem resistant Enterobacteriaceae. These organisms tend to be resistant to many antibiotics, limiting treatment options, and increasing the morbidity and mortality of the patient [1]. VRE in particular is known to be persistent, and resists many external stressors, such as chemical treatments, and can survive under a wide range of pH, temperature and salt concentrations. VRE can therefore remain on abiotic surfaces and contaminate hospital wards for many years, and is often used as an indicator to determine the cleanliness of the environment [2]. The prevalence of such nosocomial pathogens are monitored through surveillance, which encompass patient and environmental screenings, the detection, identification and strain type characterization of pathogens. Through surveillance, outbreaks are detected and actions such as patient isolation, reinforcement of hand hygiene protocol, and decontamination of wards, can be taken to interrupt further transmission to new patients [1].

Different strain typing methods can be used depending on the target pathogen, but pulsed field gel electrophoresis (PFGE) is popular and has been considered the gold standard for many bacterial pathogens, as it can be used to type many different organisms, by choosing the appropriate restriction enzyme to digest the DNA. Despite being used as the gold standard for over 20 years, PFGE has its limitations, such as the technique being laborious, requiring highly trained personnel, and has long turnaround time to results. It is not realistic to run PFGE analyses in real-time, and thus microbial strain typing is commonly conducted retrospectively [3]. Furthermore, PFGE is not suitable for long-term strain type comparison for organisms like VRE which have high recombinant rates [4]. Although the degree of relatedness between strain types are currently determined by the extent of similarity (and the number of band differences) in PFGE fragment patterns, the genetic diversity between isolates is not necessarily reflected in this way. The restriction enzyme used in PFGE cuts the genome into pieces where specific nucleotide patterns exist (i.e. CCCGGG for *SmaI* restriction enzyme), and thus depending on the location in which genetic alterations occur, through acquisition and loss of accessory genes related to pathogenicity and virulence, the extent of variation in DNA fragmentation and band patterns for PFGE can be impacted with little or vast changes, or non at all [4]. The genetic diversity amongst closely related

isolates may result in diverse PFGE band patterns, or insufficient PFGE pattern differences between unrelated isolates may occur, resulting in incorrect strain type interpretations [3]. PFGE is known to also have low data transferability, making it difficult to compare isolates characterized in different laboratories, unlike other strain typing techniques such as multilocus sequence typing (MLST) which have a standardized nomenclature based on the number and location of alleles present on the housekeeping genes [5]. Alternative strain typing options to PFGE are required where data can be acquired routinely, and provide results to infection prevention and control teams for optimized decision making. Although whole genome sequencing (WGS) is gaining popularity, and many laboratories have demonstrated successful strain typing in retrospective and prospective studies, the standardization in the method, both data acquisition and analysis is still required before it can replace PFGE as the gold standard [2, 6].

FTIR spectroscopy is a rapid, and reagent free technique that has been studied for strain typing applications [4, 7, 8]. FTIR spectroscopy has the ability to acquire spectral data within minutes, and produce results with an additional few minutes upon obtaining isolated colonies from agar plates, thereby reducing the data acquisition and analysis time compared to PFGE (Figure 5.2). With the release of Bruker's IR Biotyper, the interest in using FTIR spectroscopy for strain typing increased over the past years, and reports on subspecies typing on microorganisms such as methicillin resistant *S. aureus, Klebsiella pneumoniae,* and *Enterobacter cloacae* have been published, in ATR or transmission mode [7, 9-14]. Many of these findings approach subspecies characterization and identification similarly to how genus and species identification are achieved using FTIR spectroscopy, through the development of spectral databases based on known serotypes. In 2016, AlMasoud et al. [15] used transmission FTIR spectroscopy in combination with chemometric analyses to discriminate between strain types of 35 VRE isolates, in reference to PFGE. They reported 89 % concordance to PFGE classification, using supervised analysis (principal component analysis– linear discriminant analysis; PCA-LDA). However, PCA-LDA applied to differentiate isolates, rather than strain types, resulted in isolates clustering that was 54 % concordance to PFGE [15]. In current study, the capabilities of transflection-FTIR spectroscopy for strain typing was evaluated, using VRE with supervised and unsupervised multivariate statistical analyses. The objective was to evaluate the discriminatory capabilities and report on clustering results based on spectral differences, in reference to PFGE typing results with retrospective and prospective datasets comprised of 214 and 129 isolates respectively.

Figure 5.1 Techniques used for VRE detection and the time required for each and cumulative analyses, prior to strain type characterization at the clinical microbiology laboratory

VRE Identification

Technique	Purpose	Result	Time
VRE screening broth	Detection of *enterococci*	Black: positive	18-24 hrs max. 48 hrs
Polymerase chain reaction (PCR)	Detection of *vanA* and *vanB*	+ve -ve	4-8 hrs
Chromogenic agar	Species identification of vancomycin resistant *Enterococci*	Purple: *E. faecium* Blue-green: *E. faecalis*	24-48 hrs
Total Time:			**46 – 104 hrs**

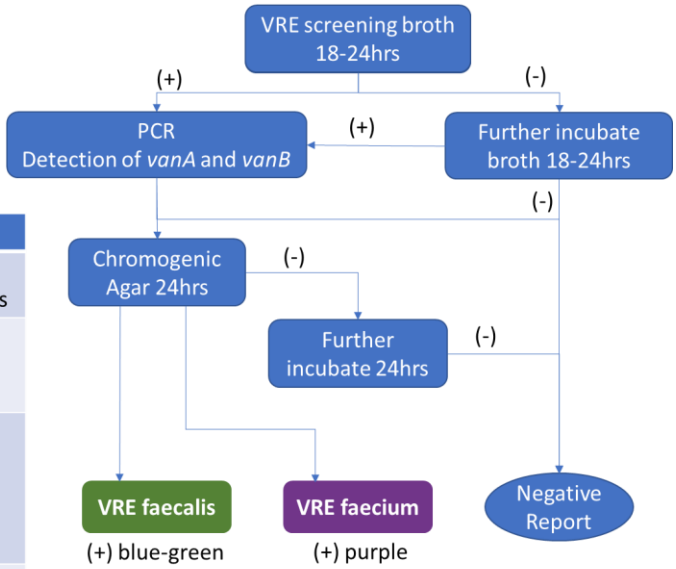
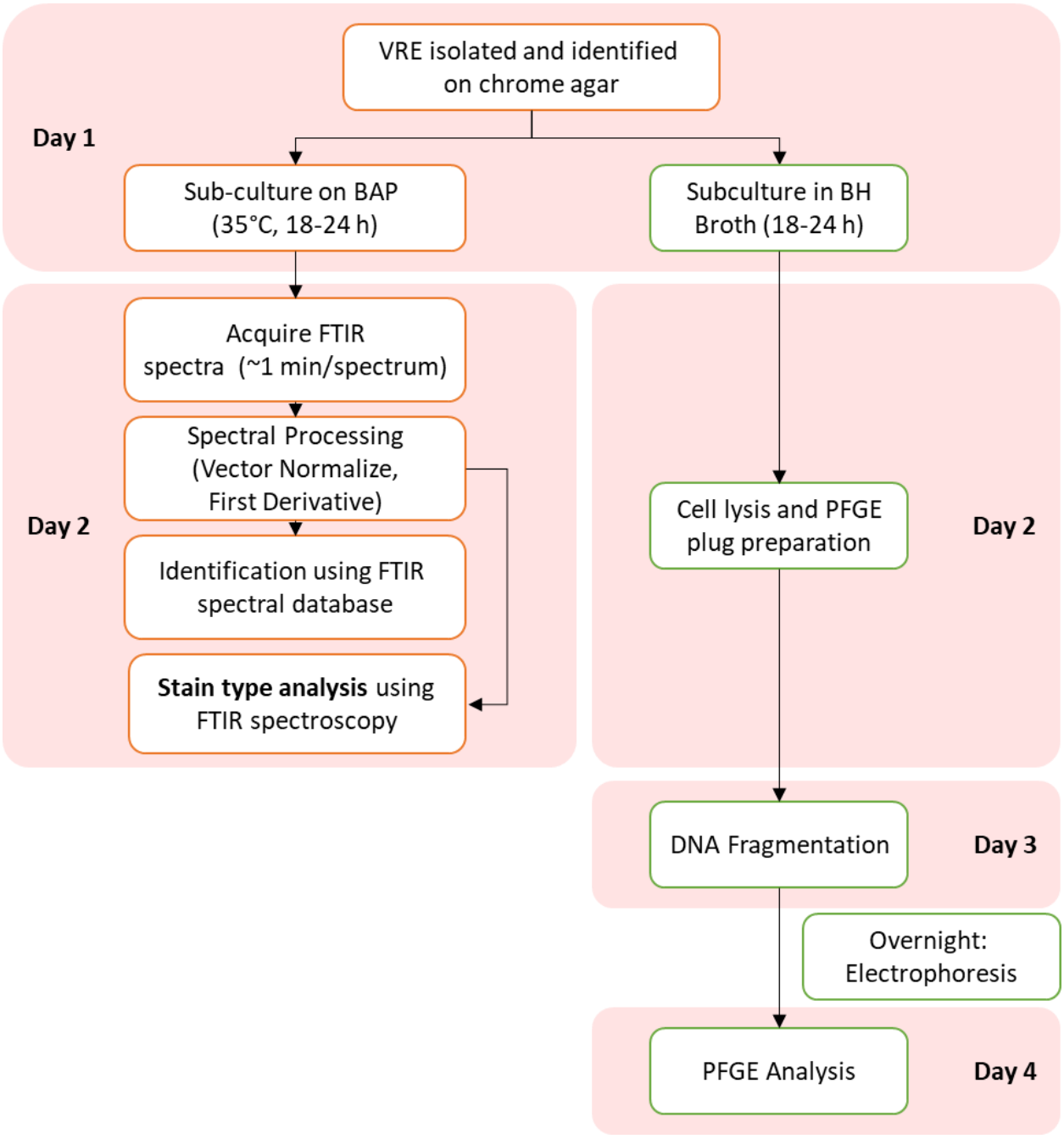


Figure 5.2 Overview of the time and steps required for strain typing by PFGE and FTIR spectroscopy, showcasing the improvement in turnaround time to results by FTIR spectroscopy.



5.3 Materials and Methods

5.3.1 Clinical isolates for retrospective and prospective VRE strain typing analysis

A set of 214 vancomycin resistant *E. faecium* (VRE) isolates, which were collected and characterized by PFGE between April 2016 and September 2017 were used in this study. The samples were obtained from patient and environmental screenings conducted at the Montreal General Hospital (). Isolates were cultured from frozen glycerol stocks onto Columbia agar with 5 % sheep blood (BAP) and incubated for 24 hours at 35 °C under aerobic condition. Prior to FTIR spectral acquisition, all isolates were sub cultured and incubated under the same conditions as the first passage, and also cultured onto VRE ChromAgar (bioMérieux SA, Marcy l'Étoile, France) for visual confirmation as VRE.

For the prospective 6-month VRE surveillance study, 129 isolates of VRE were routinely collected over a 6-month period (December 2018 to June 2019), from patient and environmental screenings at two hospitals in Montreal, Quebec, the Montreal General Hospital (MGH) and Royal Victoria Hospital (RVH). The samples were all identified as VRE using VRE screening broth (Oxoid, Neepawa, ON) , PCR (for identification of *vanA* or *vanB* gene) and VRE ChromAgar (bioMérieux SA, Marcy l'Étoile, France) (Figure 5.1). All samples were provided on VRE ChromAgar (bioMérieux SA, Marcy l'Étoile, France), and sub-cultured onto BAP (Oxoid, Neepawa, ON), and incubated at 35 °C for 24 hours prior to spectral acquisition.

5.3.2 Transflection FTIR spectral acquisition, pre-processing and analyses for microbial discrimination at the strain type level

Using sterile 1 µl disposable loops, isolated colonies were picked directly from agar plates and smeared onto IR reflective, low-E slides (Kevley Technologies, Chesterland, OH, USA) as thin films in triplicate spots and air-dried. Using the Cary 630 FTIR spectrometer (Agilent Technologies, Santa Clara, CA) with a 10° specular reflectance sample interface (Agilent Technologies, Santa Clara, CA) or a SurveyIR microscope (Czitek, Danbury, CT, USA) and MicroLab Software (Agilent Technologies, Santa Clara, CA), trasnflection spectra were acquired with 64 coadded scans at 8 cm^{-1} resolution and 8 zero-filling, in the spectral range 4000-650 cm^{-1}. Prior to spectral acquisition, background spectra were acquired from a bare surface area of the E-glass slide with the same settings as spectral acquisition. Spectra were pre-processed by taking the

first derivative and vector normalization to eliminate variability caused by baseline shifting and sample thickness respectively.

5.3.2.1 Supervised and unsupervised multivariate statistical analysis methods applied on FTIR spectra acquired from VRE isolates

First, to evaluate whether spectral differences exist between strain types, supervised and unsupervised analyses were employed on select VRE isolates. AA and CC pulsotypes (and their 1-3 band pattern variants) were chosen due to high sample representation of each types (n = 34 and 109 isolates respectively). Using the spectral region 1350-800 cm^{-1}, the spectra of these isolates were subjected to hierarchical cluster analysis (HCA) and principal component analysis (PCA) for unsupervised analysis to ascertain whether there were any global differences between the two strain types, using JMP Pro ver. 15.2.0 (Cary, North Carolina, USA). PCA was also used to reduce variables for subsequent multivariate analyses. Supervised analysis were conducted using support vector machine (PCA-SVM), and PCA-LDA, using 75 % of isolates as the training set, and remaining 25 % as validation. SVM model was made using radial basis function, with cost = 1 and gamma = 0.125. Gamma was calculated as the inverse of the number of variables.

Analysis to determine spectral relatedness between isolates were conducted with modifications to that reported by AlMasoud [15]. Briefly, variable reduction was conducted using PCA, by identifying the number of PCs that explained over 1 % of the variance among the spectra. Using these PCs, the PC scores of the isolates were subjected to PCA-LDA, using each isolate as a unique class. PCA-LDA maximized the distance between isolates, and minimized the distance between replicate spectra per isolate, by determining the weight of importance of the PCs. Squared Mahalanobis distances for each spectrum to the mean of each class, were calculated. Using the obtained squared distances, isolates were clustered by HCA with average linkage visualizing the spectral similarity between isolates in a dendrogram. Based on the squared distances observed between AA and CC pulsotype isolates, as well as between isolates within each pulsotype, a cut-off value was determined for identifying spectrally indistinguishable isolates. Spectral discrimination and clustering results were compared against PFGE pulsotype clustering, to determine the level of concordance.

5.3.2.2 Application of the developed unsupervised FTIR spectral analysis method on retrospective and prospective VRE isolates

The aforementioned PCA-LDA analysis was conducted in subsets of the dataset, divided by known pulsotype, and by unit where isolates were acquired (Table 5.1). Using the determined cut-off value (squared Mahalanobis distance < 500 as spectrally indistinguishable) in the prior analysis, the isolates were clustered based on spectral similarity. The level of concordance to PFGE pulsotype clustering were reported for each individual analysis. In the prospective study, the method was applied and analyzed for each unit/ward, where the colonized patient had stayed during hospitalization (Table 5.2) The presence or absence of clusters (an outbreak or a continuation of one) based on spectral data were determined and reported on a weekly basis, and retrospectively (2-3 weeks later) compared to PFGE results.

5.3.3 Pulsed field gel electrophoresis

PFGE analysis was conducted based on the method by Morrison et al [16]. Briefly, cells were collected from an overnight culture in Brain Heart Infusion broth, and lysed for DNA extraction using lyostaphin, lysozyme, and proteinase K. DNA was fragmented using *SmaI*. PFGE profiles were analysed using Bionumerics software (Applied Maths, Saint-Martens-Latem, Belgium). Dice coefficient with 0.5 % optimization and 1.5 % tolerance for band matching was used, and UPGMA clustering method was used to generate dendrograms in HCA.

Table 5.1 VRE isolates and their PFGE pulsotype by unit where identified

	Unit										
PFGE	11	12	13	14	15	17	18	CCU	ICU	SSU	total
A	0	0	1	0	4	0	0	0	0	0	5
A2	0	2	0	0	0	0	0	0	0	0	2
A-2	0	0	0	0	5	0	0	0	0	0	5
A-3	0	0	0	0	0	0	0	1	0	0	1
A-31	0	0	0	0	0	0	1	0	0	1	2
AA	0	8	0	1	2	2	3	1	8	0	25
AA1	0	0	0	0	0	0	1	0	0	0	1
AA-1	2	3	0	0	0	0	2	0	1	0	8
B1	8	0	0	0	0	0	0	0	1	0	9
B1-1	0	0	0	0	0	1	0	0	0	0	1
BB	0	0	0	0	0	1	0	0	0	0	1
BB-3	0	0	0	0	2	2	0	0	0	0	4
CC	6	10	0	2	18	55	5	2	5	1	104
CC1	0	0	0	0	0	1	0	0	0	0	1
CC-1	0	0	0	0	1	1	0	0	0	0	2
CC-2	0	0	0	0	1	1	0	0	0	0	2
CC-3	0	0	0	0	0	1	0	0	0	0	1
DD	0	0	0	0	0	1	0	0	0	0	1
DD-1	0	0	0	0	1	0	0	0	0	0	1
E-2	0	0	0	0	2	0	0	0	0	0	2
EE	1	0	0	0	0	0	0	0	0	0	1
FF	1	0	0	0	0	0	0	0	0	0	1
GG	0	0	0	0	0	2	0	0	0	0	2
GG-2	0	0	0	0	0	1	0	0	0	0	1
LL	0	0	0	1	1	1	0	0	0	0	3
M-2	0	0	0	1	0	0	0	0	0	0	1
M-3	0	4	0	0	0	0	0	0	0	0	4
P-2	0	0	0	0	0	0	1	0	0	0	1
Q1	0	0	0	0	0	0	0	0	1	0	1
T	0	1	0	0	0	0	0	0	0	0	1
W-1	0	1	0	0	0	0	0	0	0	0	1
X	0	0	2	0	2	1	2	0	2	0	9
X-1	0	0	0	0	0	1	0	0	0	0	1
Y	0	0	0	0	2	0	1	0	0	0	3
Z-1	0	0	0	2	1	0	1	0	0	0	4
Z-2	0	1	0	0	0	0	0	0	0	0	1
Z-3	0	1	0	0	0	0	0	0	0	0	1
Total	18	31	3	7	42	72	17	4	18	2	214

* ICU = intensive care unit, CCU = cardiac/coronary care unit, SSU = short stay unit.

Table 5.2 Hospital units and number of VRE positive patients that stayed between December 2018-May 2019

MGH			
Unit	**Number of VRE pos patients that stayed**	**Number of patients who acquired VRE in this unit**	**Number of patients who acquired VRE elsewhere**
11	2	1	1
12	7	4	3
14	1	0	1
15	12	5	7
17	6	4	2
18	4	1	3
ICU	14	5	7
RVH			
Unit	**Number of VRE positive patients that stayed**	**Number of patients who acquired VRE in this unit**	**Number of patients who acquired VRE elsewhere**
C07C	1	1	0
C07N	8	5	3
C07S	4	2	2
C07	13	8	5
C08C	5	4	1
C08N	4	3	1
C08S	12	10	2
C08			
C09C	11	7	4
C09N	1	0	1
C09S	9	9	0
C09			
C10C	4	1	3
C10N	4	3	1
C10S	3	0	3
C10	11	4	7
D03C	8	2	6
D07	3	0	3
D08N	2	1	1
D08S	2	0	2
D08	5	1	4
D09	5	2	3
D10	4	0	4

5.4 Results

5.4.1 Unsupervised and supervised analyses for FTIR spectral discrimination of AA and CC pulsotype VRE isolates

In infection control, there are two important questions to answer through strain type characterization. 1.When there is an increase in new positive cases (above the baseline) of highly pathogenic microorganisms, are these isolates indistinguishable based on microbiological tests? – Is there an outbreak starting to occur within the hospital setting? 2. Is the new positive isolate from a patient indistinguishable from isolates part of a pre-existing outbreak or endemic strains in the hospital? In previous FTIR based strain typing studies, multi-drug resistant pathogens like MRSA and KPC-positive *K. pneumoniae*, were discriminated and classified using supervised analysis and development of spectral databases, based on the capsular polysaccharide characteristics [9, 17]. *E. faecium* is typically typed for strain relatedness based on genotypic methods such as PFGE, MLST, core genome MLST and single nucleotide polymorphism (SNP) using WGS data. It is recognized that due to the high recombination rates and horizontal gene transfer in the auxiliary genome, the phenotype of *E. faecium* isolates vary extensively, even for isolates with close genetic relatedness by core genome analyses. In order to identify clonal isolates in this study, FTIR spectral data of VRE isolates were analyzed without a database, using unsupervised analyses.

AA and CC pulsotypes were the two most common strain types identified to be present over the 16 month period (April 2016-September 2017) which VRE samples were chosen from archive, with 34 and 109 isolates respectively (Table 5.1). PFGE band patterns of AA and CC pulsotypes were distinct, and show clear discrimination between them in HCA (Figure 5.3). FTIR spectra were pre-processed prior to conducting PCA on spectral region 1350-800 cm^{-1}. The first 8 principal components (PCs) contributed to 91.975 % of the total explained variable among the spectra, where each PC contributed over 1 % of the explained variable (Table 5.3). In PCA, natural clustering of AA and CC isolates were observed in a biplot of PC 2 vs PC 4, indicating that there are global spectral differences between the two pulsotypes (Figure 5.4).

SVM model was developed using 75 % of each pulsotypes, with the 8 PCs. The remaining 25 % of isolates were used as a validation set, to determine if the supervised discrimination model can correctly predict AA and CC pulsotypes. The discrimination between AA and CC pulsotypes were successful at 97.2 % concordance, with 4 outliers identified (3 and 1 isolates from the training

and validation sets respectively). The isolates that were correctly identified were predicted with a mean probability of 93.6 % (confidence interval (CI) 95 % [94.3-92.9 %]) and 98.8 % (CI 95 % [99.0-98.7 %]) as AA and CC respectively, indicating high confidence for these predictions. Among the 4 outliers, one outlier predicted as CC (identified as AA by PFGE) with > 96 % probability on 5 of 6 spectral replicates (6^{th} spectrum had 85 % probability as CC (Table 5.4)). The remaining three outliers were isolates identified as CC by PFGE that were spectrally more similar to AA isolates and were predicted as AA with probabilities ranging between 75 and 90 %. Upon re-analyzing the outliers by PFGE, it was confirmed that the original PFGE strain type identification for three of the four isolates was incorrect, and the results of the PFGE re-analysis for these three isolates was concordant with the FTIR results (Table 5.4). In addition to supervised analysis employing PCA-SVM, unsupervised analyses of the spectra of the VRE isolates were conducted by PCA-LDA and HCA. In PCA-LDA, the spectral replicates acquired for each isolate were labeled as an individual group (by sample ID), and a model was created that maximizes distances between groups and minimizing distances between spectral replicates. Averaged squared Mahalanobis distances were calculated between all isolates subjected in the PCA-LDA. These values were then used in HCA, with average linkage to show relatedness (Figure 5.5). The HCA was in agreement with the PCA-SVM analysis in regards to the 4 observed outliers.

A total of 15 isolates that were variants of the AA and CC pulsotypes (9 and 6 respectively) were included in the study. Variants were of the pulsotypes were given labels such as -1 and -2, which indicate the presence or loss of one or two band features. Isolates that are variants of AA and CC pulsotype are not as distinctive in the HCA based on spectral data from non-variant AA and CC pulsotype isolates. This indicates that the spectral features do not necessarily reflect the genotypic changes observed in PFGE. In some cases, the genetic evolution or change is reflected in spectral changes. For example, CC1 isolate (dull light green highlighted in Figure 5.5), which had an extra band in the PFGE band pattern relative to the CC pulsotype (and 76.9 % similarity to other CC isolates by PFGE), had a higher spectral similarity to AA isolates compared to CC isolates. In the dendrogram obtained by HCA, it was observed to fall under the main arm that clusters AA isolates together. In the SVM analysis, when the CC1 isolates was removed from the training set, it had 42.2 % and 57.8 % probability of being AA and CC respectively, indicating that it is not spectrally similar to either clusters. Although this isolate was re-analyzed by both PFGE and FTIR spectroscopy, the discordant results remained. A CC isolate (XIV315, most right isolate

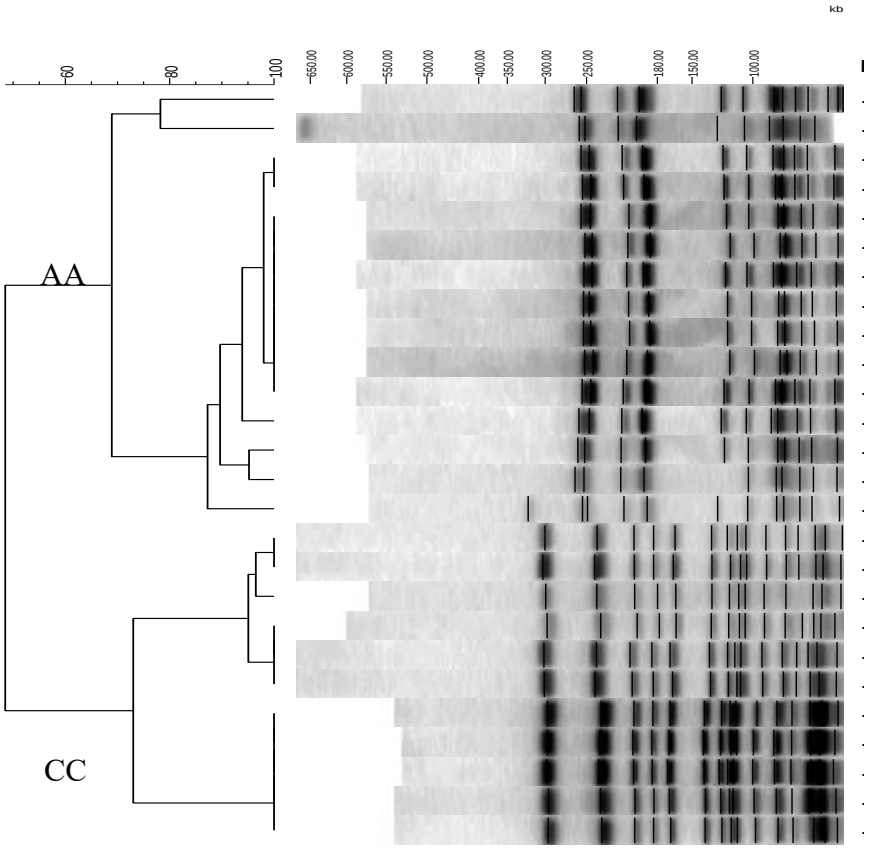
(Figure 5.5) based on the HCA was not spectrally similar to all AA isolates and most CC isolates. Based on the spectral analyses this isolate could not be considered a CC pulsotype. Since this isolate was used as part of the training set in SVM analysis, it was not identified to be an outlier, and had high probability of being CC. When removed from the training set, this isolate had a probability of 14.2 % and 85.8 % as AA and CC respectively. Considering that other isolates had probability values ranging between the mid and high 90s, the probability reflects the lesser degree of spectral similarity between this isolate to other CC isolates.

The distinction between clusters was also visualized by plotting the squared Mahalanobis distances of an isolate to all other isolates (Figure 5.6). The plots were colored based on the pulsotype designation provided by PFGE. The clustering of isolates corresponded with the squared distances. For instance, in Figure 5.6, it was evident that the sample being analyzed (XVIII822) was an AA, due to its small squared distances to other AA isolates (where smaller squared distance means higher spectral similarity). These squared distances were less than 500, while some outliers, and variants of AA were observed to have a squared distance of 600 or less. Spectral similarity to non-AA isolates was also observed, and these isolates were the outliers observed in both the SVM and HCA results. The scatterplot also visualized the distinct spectral differences between AA and CC pulsotypes. Three outliers that were previously identified, also show high spectral similarity to this isolate (68-400 squared Mahalanobis distance), whereas the remaining CC isolates had squared distances mostly over 1100, and an average of 2746. The AA outlier was also 2647 squared distance away, indicating that it is clustering with CC, and not with other isolates that were identified as AA pulsotype. Based on pre-identified PFGE pulsotypes (AA and CC) and the squared distances calculated between these isolates, cut-off values were determined (Table 5.5). Isolates were considered spectrally indistinguishable when the squared distance was less than 500, while isolates were considered subclusters when squared distance was between 500 and 1000. Isolates with squared distances over 1000 were determined as not related. AA and CC isolates were spectrally clustered into two main groups, with two subclusters (one per group), and three singleton isolates, that were spectrally unrelated to all other isolates (Table 5.6). For AA and CC pulsotype VRE isolates, these cut-off values provided high concordance to PFGE strain typing, with 91.7 % and 91.0 % categorical agreement respectively.

Cluster 2 and 4 were sub-clusters that are spectrally similar to some but not all isolates that belong in Cluster 1 and 3 respectively. Based on the squared distances and the cut-offs, Cluster 1

and 3 represent isolates that were identified as AA and CC respectively. When comparing to PFGE pulsotypes, the true positive rates for AA was 88 %, and CC was 90.4 %. False positive rates for AA and CC were 9.2 % and 15 % respectively, mostly as a result of PFGE variants clustering with the main clusters(Table 5.7). Most of the discrepancies between FTIR spectral based and PFGE typing arose from the clustering of variant PFGE types with the corresponding non-variants in FTIR based clustering (Table 5.6). In the following sections, the method was applied to evaluate whether clusters of isolates associated with outbreaks can be identified, in retrospective analysis by unit and pulsotype, using 214 isolates, as well as in a 6-month prospective study.

Figure 5.3 Dendrograms obtained by HCA of PFGE pattern data for sub-set of VRE isolates, AA and CC.



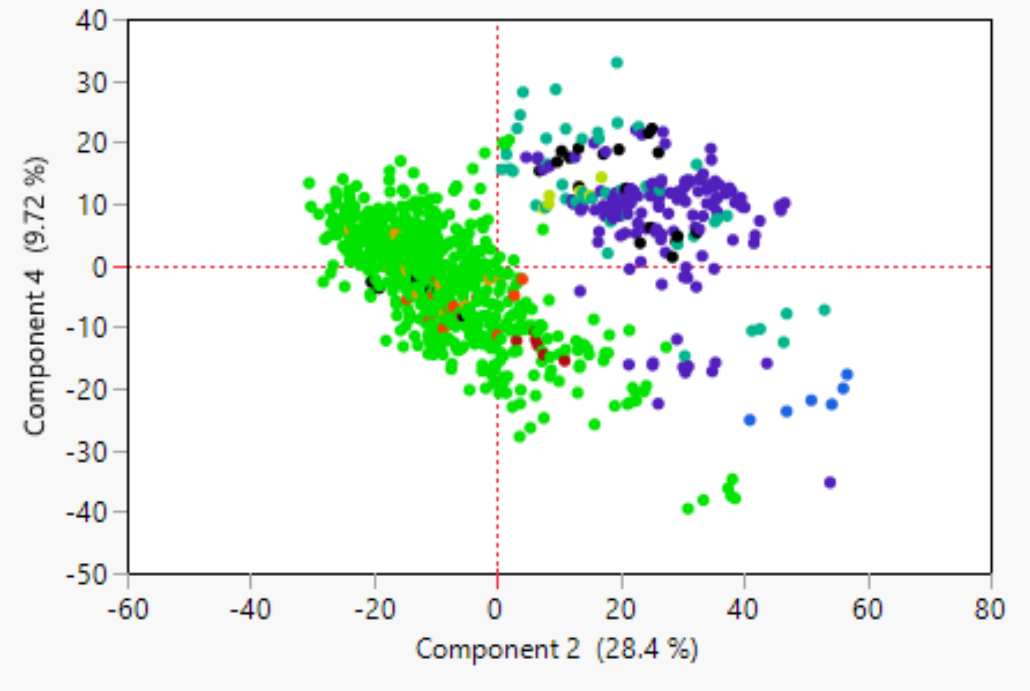
Clear differentiation between CC and AA strains (sub-set of all isolates used in this analysis) observed in the PFGE pulsotype.

Table 5.3 Principal components and explained variable (%) for AA and CC VRE spectra

PC	Eigenvalue	Explained Variable (%)	Cumulative Explained Variable (%)
1	367.5065	31.118	31.118
2	335.3489	28.395	59.514
3	166.0014	14.056	73.570
4	114.7855	9.719	83.289
5	33.7482	2.858	86.147
6	29.9060	2.532	88.679
7	25.8306	2.187	90.866
8	13.1311	1.112	91.978
9	9.7870	0.829	92.807
10	7.8444	0.664	93.471

The principal components, and their respective eigenvalue and percent variable explained within spectra of VRE isolates of AA and CC pulsotypes.

Figure 5.4 PCA in spectral region 1350-800 cm^{-1}, showcasing global spectral differences between VRE AA (purple) and CC (green) pulsotypes in score plot for PC 2 vs PC 4.



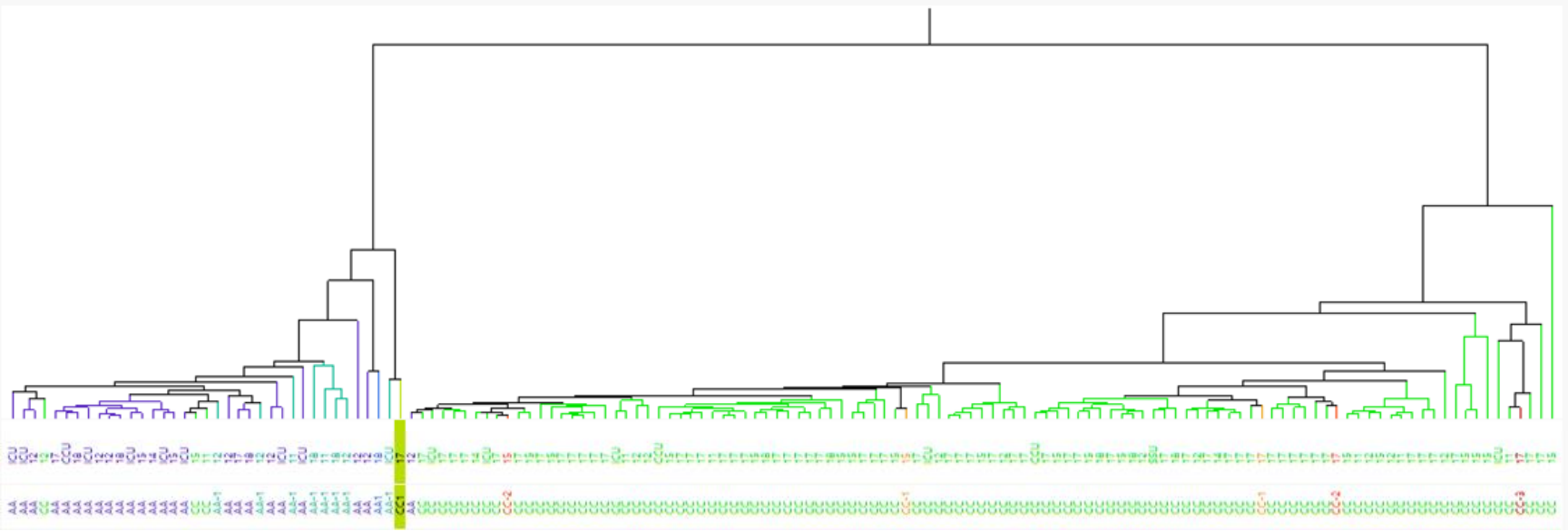
PCA in broad fingerprint spectral region showed global differences exist between AA (purple) and CC (green) isolates.

Table 5.4 List of outlier isolates identified by SVM and typing results from PFGE

Outliers	Training / Validation	Unit Tested	Date Acquired	PFGE Pulsotype	FTIR (PC-SVM)	Averaged probability by PC-SVM as AA	Averaged probability by PC-SVM as CC	PFGE 2nd analysis
XVII476	Training	12	Jan 2017	AA	CC	4.6%	95.4%	CC
XVII347	Validation	11	May 2017	CC	AA	89.1%	10.9%	AA
XVII639	Training	15	May 2017	CC	AA	85.6%	14.4%	AA
XVII936	Training	12	Jan 2017	CC	AA	75.7%	24.3%	N/A
XVII456	Training / Validation	17	Sept 2016	CC1	CC / AA	2.6% 42.2%	97.4% 57.8%	CC
XVII315	Validation	15	July 2017	CC	CC	14.2%	85.8%	N/A

Outliers from the SVM analysis, as well as an outlier observed from HCA (Figure 5.5). PFGE analysis was not redone on sample XVII315 and XVII639.

Figure 5.5 Dendrogram of averaged replicate spectra of AA and CC isolate showcasing spectral differences between the two strain types in an unsupervised HCA



The dendrogram was generated using average linkage and Euclidean distance. The HCA showcased global clustering of CC isolates (green) and AA isolates (purple) .The HCA also was in agreement with the SVM analysis in regards to the 4 observed outliers. CC-2 (red), CC-1 (Orange) clustered with CC isolates, and were considered spectrally indistinguishable from CC isolates. AA-1 (turquoise) and AA1 (blue) clustered with AA isolates.

Figure 5.6 Averaged, squared Mahalanobios distance to each isolate for sample XVIII822

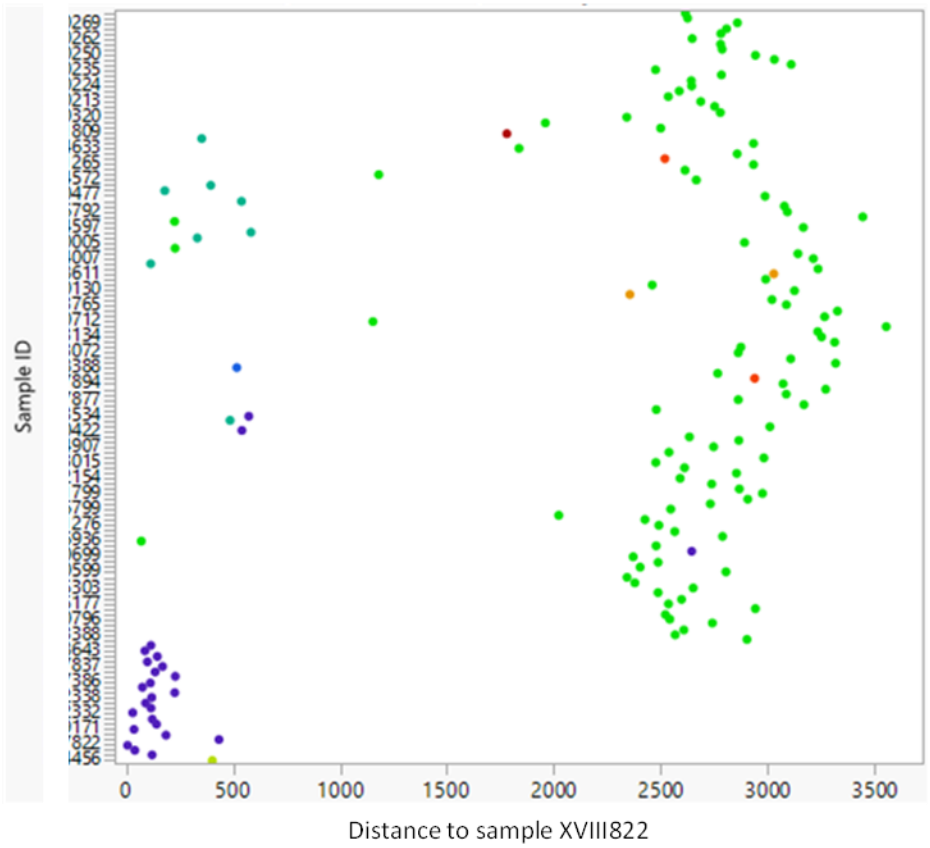


Table 5.5 Cut-off value determined for FTIR spectroscopy based on strain type difference AA and CC VRE pulsotypes

Squared distance	**FTIR spectral result**
0-500	Clonal/cluster
500-1000	Potential spectral similarity
> 1000	Not spectrally considered clonal / part of a cluster

Table 5.6 Clusters identified based on FTIR spectral analyses of AA and CC VRE pulsotypes

PFGE	Isolates	Cluster 1	Cluster 2*	Cluster 3	Cluster 4*	Cluster 5	Cluster 6	Cluster 7
AA	26	24	2	0	0	0	0	0
AA1	1	0	1	0	0	0	0	0
AA-1	8	8	0	0	0	0	0	0
CC	103	1	0	95	4	1	1	1
CC1	1	1	0	0	0	0	0	0
CC-1	2	0	0	2	0	0	0	0
CC-2	2	0	0	2	0	0	0	0
CC-3	1	0	0	1	0	0	0	0
Total	144	34	3	100	4	1	1	1

Cluster 2 and 4 were sub-clusters that are spectrally similar to some but not all isolates that belong in Cluster 1 and 3 respectively. Based on the squared distances and the cut-offs, Cluster 1 and 3 represent isolates that were identified as AA and CC respectively. When comparing to PFGE pulsotypes, the true positive rates for AA was 88 %, and CC was 90.4 %. False positive rates for AA and CC were 9.2 % and 15% respectively, mostly as a result of PFGE variants clustering with the main clusters.

Table 5.7 Concordance between PFGE and FTIR spectral typing results for clustering of AA and CC pulsotype VRE isolates

	AA	**AA1**	**CC**
True Positive	92.3 %	100 %	92.2 %
False Positive	8.5 %	1.4 %	12.2 %
True Negative	91.5 %	98.6 %	87.8 %
False Negative	7.7 %	0 %	7.8 %
Categorical agreement	91.7 %	98.6 %	91.0 %

Based on the squared Mahalanobis distances calculated between AA and CC pulsotype isolates in PC-LDA, the clustering of isolates based on spectral similarity resulted in 91.7 % and 91.0 % categorical agreement to PFGE for AA and CC pulsotype isolates respectively.

5.4.2 Unsupervised analyses for simulation of prospective VRE typing by FTIR spectroscopy

Following the unsupervised, PCA-LDA method described in the previous section, a total of 214 VRE samples (191 patient screening and 23 environmental samples) collected between April 2016 and July 2017 were analysed for the spectral relatedness. The samples were previously identified by PCR and VRE ChromAgar (bioMérieux SA, Marcy l'Étoile, France) as vancomycin resistant *E. faecium*, and the strain relatedness between isolates were determined by PFGE. The samples were prepared from frozen (-80°C), onto Columbia agar with 5 % sheep blood (BAP) and reconfirmed as VRE using Chrome agar. PCA was conducted on all spectra from 214 isolates, identifying 9 PCs that each contribute > 1 % of total variance between spectra. These 9 PCs were subsequently used in LDA, as variable reduction The PCA-LDA was applied to the entire dataset, as well as subsets of VRE isolates, divided by PFGE pulsotypes, and hospital units where patients were screened.

In Table 5.8 the results of analyses conducted per pulsotype and the relative concordance between FTIR spectral typing results and PFGE pulsotypes are presented. For pulsotypes that had large numbers of isolates present like AA (n=24) and CC (n=101), the concordance between PFGE and FTIR analyses was > 90 % for true positive rate, indicating that isolates that were identified to be spectrally indistinguishable were also identified to be clonal by PFGE. Among the isolates classified by PFGE as AA and CC variants, some clustered spectrally with AA and CC isolates (and hence were false positives), while some were part of subclusters, or determined to be unrelated to isolates of their corresponding non-variant pulsotype. Although in PFGE analyses, isolates with 1-3 band difference are considered closely or possibly related [3], the FTIR typing results indicate that not all isolates with the same PFGE band variation (ie AA-1) have the same phenotype, and can cluster separately from AA and remaining AA-1 isolates (which had no observable spectral difference from AA isolates). In the case of CC pulsotype variants CC-1, CC-2 and CC-3, four out of five isolates were spectrally indistinguishable from CC pulsotype isolates, and therefore despite PFGE band differences (thereby also genetic changes), these changes had no effect in the FTIR spectral features of the VRE isolates. Similar non-discriminatory (and hence false positive) results were observed for Z-3 pulsotype isolate that was spectrally similar to the Z-1 isolates. For pulsotype GG, none of the three isolates were considered spectrally clonal. However, upon closer inspection of the PFGE results, the three isolates had inconsistent PFGE band patterns, despite given the same pulsotype. Such inconsistent naming system caused by manual error can

occur, within a single laboratory by one person (as observed in this example), and demonstrates the difficulty to standardize PFGE results for comparison. Reproducibility between PFGE runs can occur, further complicating the pairwise comparisons [3]. Similarly, BB-3 pulsotype isolates were analysed with a BB pulsotype isolate, and resulted in two distinct clusters, each comprised of two isolates, with the fifth isolate being sporadic and spectrally not related to either clusters. Overall, the FTIR spectrotyping method resulted in 67-100 % true positive rates relative to PFGE pulsotypes. For four pulsotypes, LL (n = 3), M-3 (n = 4), X (n = 10), and Z-1 (n = 4), the true positive rates were 100 %, showcasing that clonal isolates cluster together based on spectral similarity, despite the fact that PCA-LDA is modeled to maximize distances between all isolates. Although the categorical agreement upon analysis with variants are important, as it describes the FTIR spectrotyping's ability to reliably differentiate sporadic isolates from a cluster, the effect on the spectral patterns and therefore clustering is not consistent enough to consider these variant isolates as separate nor part of the clusters. In routine analyses, the strain types are yet to be known, and therefore the most important factor is whether accurate identification of clonal isolates can be achieved or not. Based on the results in Table 5.8, FTIR spectrotyping method demonstrated its ability to identify isolates that are indistinguishable (and thus considered "clonal" to each other) based on spectral features.

In the analyses conducted by hospital unit, the method was tested to determine if it was capable of discriminating between various unrelated pulsotypes, while clustering clonal isolates. Table 5.1 shows the prevalence of each pulsotype by units, and the corresponding results from FTIR spectral typing in reference to PFGE pulsotypes are presented in Table 5.11. Overall, clustering was observed by FTIR spectrotyping in Units 11, 12, 13, 14, 18, CCU, ICU and SSU. Categorical agreement was calculated for each pulsotype with more than one isolate represented in the unit-by-unit analysis, and ranged from 26.8 % to 100 %. Low concordance between PFGE and FTIR spectrotyping was observed when isolates with various pulsotypes were considered clonal and clustered together. This was the case for a cluster in Unit 15 (Table 5.7). FTIR spectrotyping of 41 isolates from Unit 15 yielded 3 clusters and 3 sporadic isolates whereas these were clustered into 11 pulsotypes by the PFGE analysis, of which 3 were considered sporadic. Only one of the three sporadic isolates were in agreement between FTIR spectrotyping and PFGE pulsotype. Meanwhile, the two other sporadic isolates by PFGE (DD-1 and CC-2) were respectively clustered with isolates that were predominantly AA and CC pulsotype. In Unit 11,

there were 15 VRE positive patients identified between December 2016 and August 2017. Seven of these were isolated in January 2017, prompting an environmental screening for VRE, where three additional samples were identified. Based on the PFGE results, there was an outbreak caused by pulsotype B1, with two other potential clusters occurring, unrelated to the outbreak. However, based on FTIR spectrotyping, there were two outbreaks, one outbreak constituting of isolates that were designated pulsotypes B1 and AA-1, and another cluster composed of CC pulsotypes. The sporadic isolate identified by PFGE (EE) was also identified as a sporadic isolate by FTIR spectrotyping. FF pulsotype isolate however clustered closely with the CC pulsotype isolates by FTIR spectrotyping. Despite the varying categorical agreement results, the true positive rates for each pulsotype ranged from 80 % to 100 %. The method has the ability to accurately identify isolates with the same PFGE pulsotypes as spectrally indistinguishable. However, the high false positive rates in some cases indicate either a lack of spectral difference between pulsotypes, despite the PFGE differences, or a need for improvement in the cut-off value for clonal determination in the method. The developed method was further tested in a 6-month prospective study, to identify clonal isolates and outbreaks.

Table 5.8 Comparison of FTIR spectrotyping and PFGE results by pulsotypes of retrospective VRE isolates

PFGE Pulsotype	No. of isolates per pulssotype	Cluster	True positive rate	False negative rate	False positive rate	True negative rate	Categoric agreement	Subcluster 1	Subcluster 2
AA	24	22	91.7 %	8.3 %	55.6 %	44.4 %	78.8 %	2	0
AA1	1	0						1	0
AA-1	8	5						0	3
B1	8	6	75 %	25 %	-	-	75 %	2	
BB	1	2	50 %	50 %	0 %	100 %	60 %	2	
BB-3*	4								
CC	101	93	92.1 %	7.9 %	80 %	20 %	88.7 %	4	
CC-1	2	1							
CC-2	2	2							
CC-3	1	1							
DD	1	No					100 %		
DD-1	1	cluster							
GG	2	No					0 %		
GG-2	1	cluster. All non-related							
LL	3	3	100 %	0			100 %		
M-3	4	4	100 %	0			100 %		
X	10	10	100 %	0			100 %		
Y	3	2	66.7 %	33.3 %			66.7 %		
Z-1	4	4	100 %	0	50 %	50 %	83.3 %		
Z-2	1	0							
Z-3	1	1							

The 5 BB isolates spectrally separated into two clusters of two isolates, and one unrelated isolate.

Table 5.9 Clusters of isolates identified by FTIR spectrotyping per unit in the retrospective VRE isolates

Unit	Clusters **observed** by spectrotyping	**PFGE pulsotype**	**True Positive**		**False Positive**		**True Negative**		**False Negative**		**Categorical Agreement**	
11	Cluster 1: B1 (8) /AA-1 (2) / FF (1) Cluster 2: CC (5) 2 sporadic isolates: EE (1) / CC (1)											
	18	B1	8	100 %	3	30 %	7	70%	18	B1	8	100 %
		AA-1	2	100 %	9	56.3 %	7	43.8%	0	AA-1	2	100 %
		CC	5	83.3 %	1	8.3 %	11	91.7%	1	CC	5	83.3 %
12	Cluster 1: CC (10) /A2 (2) / T (1) / Z-2 (1) / Z-3 (1) Cluster 2: AA (7) / AA-1 (3) / CC (1) / W-1 (1) / M-3* (4) *M-3 observed as a subcluster of AA cluster											
	31	CC	10	90.9 %	5	25 %	15	75%	31	CC	10	90.9 %
		A2	2	100 %	13	44.8 %	16	55.2%	0	A2	2	100 %
		AA	7	100 %	9	36 %	15	62.5%	0	AA	7	100 %
		AA-1	3	100 %	13	46.4 %	15	53.6%	0	AA-1	3	100 %
		M-3	4	100 %	12	44.4 %	15	55.6%	0	M-3	4	100 %
13	Cluster: X (2) A was sporadic											
	3	X	2	100 %	0	0 %	1	100%	3	X	2	100 %
14	Cluster 1: AA (1) /M-2 (1) Cluster 2: CC (2) /Z-1 (2) LL is sporadic											
	7	AA/M-	0	0 %	2	33.4 %	5	83.3%	7	AA/M-	0	0 %
		CC	2	100 %	2	40 %	3	60%	0	CC	2	100 %
		Z-1	2	100 %	2	40 %	3	60%	0	Z-1	2	100 %

(*continued*)

Unit	Isolates observed per unit	PFGE pulsotype	True Positive		False Positive		True Negative		False Negative		Categorical Agreement	
15	Cluster 1: CC (16) /A-2 (5)/ E-2 (2) /Z-1 (1) / A (4) / BB-3 (2) / Y (1) / CC-2 (1), Cluster 2: AA (3) / DD-1 (1) Cluster 3: X (2) 3 sporadic isolates (Y, CC, LL)											
	41	CC	16	94.1 %	16	66.7 %	8	33.3 %	41	CC	16	94.1 %
		A-2	5	100 %	27	75 %	9	25 %	0	A-2	5	100 %
		E-2	2	100 %	30	76.9 %	9	23.1 %	0	E-2	2	100 %
		A	4	100 %	28	75.7 %	9	24.3 %	0	A	4	100 %
		BB-3	2	100 %	30	76.9 %	9	23.1 %	0	BB-3	2	100 %
		AA	3	100 %	1	2.6 %	37	97.4 %	0	AA	3	100 %
		X	2	100 %	0	0 %	39	100 %	0	X	2	100 %
17	Cluster 1: CC (51) / BB-3 (2) / CC-1(1) / CC-3 (1) / GG (1) / LL (1) Cluster 2: CC (3) / CC-2 (1) Cluster 3: GG (1) / GG-2 (1) Cluster 4: X (1) / X-1 (1) / BB (1) one CC isolate was sporadic											
	66	AA	2	100 %	3	4.3 %	67	95.7 %	66	AA	2	100 %
		CC	51	92.7 %	6	35.3 %	11	64.7 %	4	CC	51	92.7 %
18	Cluster 1: CC (5) / Y (1) / Z-1 (1) Cluster 2: AA (3) / AA-1 (1) / X (1), Cluster 3: AA1 (1) / AA-1 (1) / X (1) subgroup of AA, 2 sporadic isolates P-2 and A-31											
	17	CC	5	100 %	2	16.7 %	10	83.3 %	17	CC	5	100 %
		AA	3	100 %	2	14.3 %	12	85.7 %	0	AA	3	100 %

(*continued*)

Unit	Isolates observed per unit	PFGE pulsotype	True Positive		False Positive		True Negative		False Negative		Categorical Agreement	
CCU	Cluster: CC (2) Sporadic isolates A-3 and AA											
	4	CC	2	100 %	0	0 %	2	100 %	4	CC	2	100 %
ICU	Cluster 1: AA (7) Cluster 2: AA(1) AA-1(1), Q1(1) Cluster 3: CC(4). Cluster 4: X(2) B1(1) One sporadic isolate: CC											
	18	AA	7	87.5 %	0	0 %	10	100 %	18	AA	7	87.5 %
		CC	4	80. %	0	0 %	13	100 %	1	CC	4	80. %
		X	2	100 %	1	6.3 %	15	93.8 %	0	X	2	100 %
		AA-1/Q1/AA			3	16.7 %	15	83.3 %	0	AA-1/Q1/AA		
SSU	Two sporadic isolates (no clusters)											**100 %**

5.4.3 Six-month prospective analyses for VRE outbreaks

In this prospective study, the VRE isolates collected from patients staying at the Montreal General Hospital and Royal Victoria Hospital were analyzed by the units where patients were hospitalized, to determine whether the FTIR spectrotyping method has the ability to identify clonal isolates routinely for surveillance purposes. Taking into consideration that patients often move between wards, patients that moved between wards were analyzed and compared to other patients' VRE isolates by date that patients were admitted into the ward or when VRE screening results were positive. By combining the epidemiological data and spectrotyping results, the potential routes of VRE transmission between wards can be speculated. Twenty-five patients at Montreal General Hospital moved between 2 to 8 units (median = 4), and 2 to 20 rooms (median = 8) during their stay (0-204 days, median = 14 days). Ten of these patients were admitted into the hospital more than once (2 – 5 times) during the study period. One patient was a previously known VRE carrier, while another patient was identified as a new VRE carrier. VRE isolates were collected from 64 patients at Royal Victoria Hospital, where patients stayed in 1-10 wards (median = 5), and 1-27 rooms (median = 9), over a period of 0-218 days (median = 17 days).

Out of 16 analyses (one per ward) that were conducted with two VRE isolates, 6 analyses (37.5 %) had discordant results with PFGE. Out of these 6 wards, 4 had isolates that were not clonal, while 2 units had isolates that were considered clonal by PFGE. When the analyses were conducted with 3 samples, concordance between FTIR spectrotyping and PFGE was improved, with only 1 ward (6.7 %) with discordant results, and 3 wards (20 %) that had partially concordant clustering. In two of the three cases, isolates that were considered to have 1 or 3 PFGE band differences (e.g. BB and BB-3, QQ and QQ-1) were determined to be spectrally clonal. In the last case, the third isolate that was analysed was spectrally clonal to one of the two pre-existing isolates, despite all three isolates having the same pulsotype designation.

Table 5.10 FTIR spectrotyping results for clusters of isolates for potential outbreak detection, in comparison with PFGE pulsotype designation for MGH units at the end of the study period.

Ward	Isolates	PFGE cluster designation	True Positive (n, %)		False Positive (n, %)		True Negative (n, %)		False Negative (n, %)		Categorical Agreement (n, %)
12	5	Z-1	3	100.0 %	1	50.0 %	1	50.0 %	0	0.0 %	80.0 %
15	7	AAA	2	50.0 %	0	0.0 %	5	100.0 %	2	50.0 %	71.4 %
15	7	AAA1	2	100.0 %	0	0.0 %	5	100.0 %	0	0.0 %	100.0 %
17	4	AAA	2	100.0 %	0	0.0 %	2	100.0 %	0	0.0 %	100.0 %
17	4	BB/BB-3	-	-	2	50.0 %	2	50.0 %	-	-	50.0 %
18	2	Z-1/BB	-	-	2	100.0 %	0	0.0 %	-	-	0.0 %
ICU	6	AAA	4	100.0 %	0	0.0 %	2	100.0 %	0	0.0 %	100.0 %
ICU	6	BB/Z-1	-	-	2	33.3 %	4	66.7 %	-	-	66.7 %
all	18	BB/Z-1	8	100.0 %	1	10.0 %	9	90.0 %	0	0.0 %	94.4 %
all	18	BB	3	100.0 %	6	40.0 %	9	60.0 %	0	0.0 %	66.7 %
all	18	Z-1	5	100.0 %	4	30.8 %	9	69.2 %	0	0.0 %	77.8 %
all	18	AAA	4	80.0 %	1	7.7 %	13	92.3 %	1	20.0 %	88.9 %
all	18	AAA1	2	100.0 %	0	0.0 %	16	100.0 %	0	0.0 %	100.0 %

Table 5.11 FTIR spectrotyping results for clusters of isolates for potential outbreak detection, in comparison with PFGE pulsotype designation for RVH wards at the end of the study period

Ward	Isolates	PFGE cluster designation	True Positive (n, %)		False Positive (n, %)		True Negative (n, %)		False Negative (n, %)		Categorical Agreement (n, %)
C07N	7	V-2	4	100.0 %	0	0.0 %	3	100.0 %	0	0.0 %	100.0 %
C07S	3	V-2	2	100.0 %	0	0.0 %	1	100.0 %	0	0.0 %	100.0 %
C08C	5	BBB	2	100.0 %	0	0.0 %	3	100.0 %	0	0.0 %	100.0 %
C08N	4	T	3	100.0 %	0	0.0 %	1	100.0 %	0	0.0 %	100.0 %
C08S	12	T	10	100.0 %	1	33.3 %	2	66.7 %	0	0.0 %	92.3 %
C09C	12	HH-1	2	66.7 %	0	0.0 %	9	100.0 %	1	33.3 %	91.7 %
C09S	10	CC	3	100.0 %	3	42.9 %	4	57.1 %	0	0.0 %	70.0 %
C09S	10	QQ/QQ-1	-	-	2	20.0 %	8	80.0 %	-	-	80.0 %
C10	9	BBB	2	100.0 %	1	14.3 %	6	85.7 %	0	0.0 %	88.9 %
C10	9	QQ-2 /EEE/HH/ unidentified	-	-	4	44.4 %	5	55.6 %	-	-	55.6 %
D03	7	BBB	4	100.0 %	1	25.0 %	3	75.0 %	0	0.0 %	85.7 %
D03	7	HH/V-2	-	-		28.6 %		71.4 %	-	-	71.4 %
D07	3	T/QQ-2/ unidentified	-	-	0	0.0 %	3	100.0 %	-	-	100.0 %
D09	4	BBB/T	-	-	2	50.0 %	2	50.0 %	-	-	50.0 %
D10	4	T	2	100.0 %	0	0.0 %	2	100.0 %	0	0.0 %	100.0 %

FTIR spectrotyping identified 8 clusters in 5 wards at MGH (Table 5.8). In unit 15, 17 and ICU, two clusters were identified per ward. Out of the 8 clusters, 5 clusters were concordant to the PFGE results, with 71-100 % categorical agreement. Three clusters identified by FTIR spectrotyping were not identified as clusters by PFGE. Two of these clusters contained isolates with pulsotype Z-1 and BB (in Unit 18 and ICU). Based on an analysis that was conducted on all samples from MGH, it was revealed that all BB, BB-3 and Z-1 isolates clustered together by FTIR spectrotyping. Based on the PFGE band patterns, the two pulsotypes are 84 % similar (Figure 5.7). Isolates of AAA pulsotype, another cluster of isolates that were identified within MGH wards, varied in PFGE pulsotype by up to 23 % (minimum 77 % similarity) within the same pulsotype (Figure 5.8) This indicated that Z-1 and BB pulsotypes are more similar by PFGE, than some AAA pulsotype isolates are to each other; this coincides with FTIR spectrotyping results. Despite examples of isolates belonging to variants of a given pulsotype having spectral clonality to their non-variant type, not all variants are spectrally similar to their non-variant type. This was observed in Unit 15, where AAA1 pulsotype isolates, were spectrally distinct from AAA isolates, resulting in two separate clusters, rather than one.

Analyses revealed that isolates that are spectrally similar were identified in 4 out of 5 units. For example, pulsotype AAA isolate colonized 3 patients during their stay in ICU (n = 2) and Unit 15 (n = 1). These patients and one more patient who had the AAA puslotype VRE stayed in 4 units. Based on the dates that the patients stayed in each ward, the two patients that acquired VRE in ICU were not overlapping in stay. Although frequent change in wards where patients stayed were observed, the likelihood that the same strain was transmitted across multiple wards in the hospital was low, based on their movement and timeline. The patient who stayed in the ward after the first patient was VRE positive was already VRE positive prior to staying in the said ward, indicating that the transmission of VRE likely occurred via other routes. The main source of the transmission to patients seem to be from the environment, as patient to patient transmission seems unlikely due to lack of overlap in stay, as well as the dates and location where patients are suspected to have acquired VRE. Persistent colonization of VRE in the environment is likely the cause, and this may explain why patients from different units acquired the same strain type VRE.

Figure 5.7 Pulsotype BB and Z-1 for VRE isolates in HCA based on PFGE band patterns

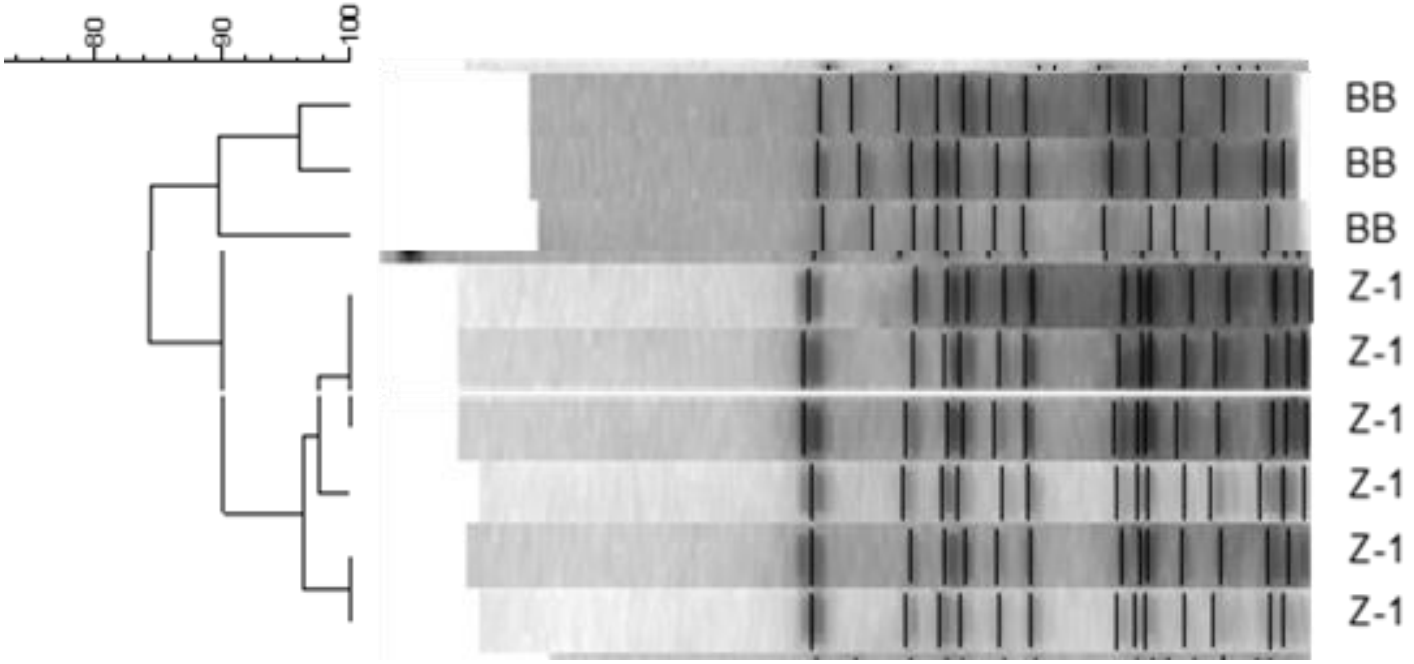
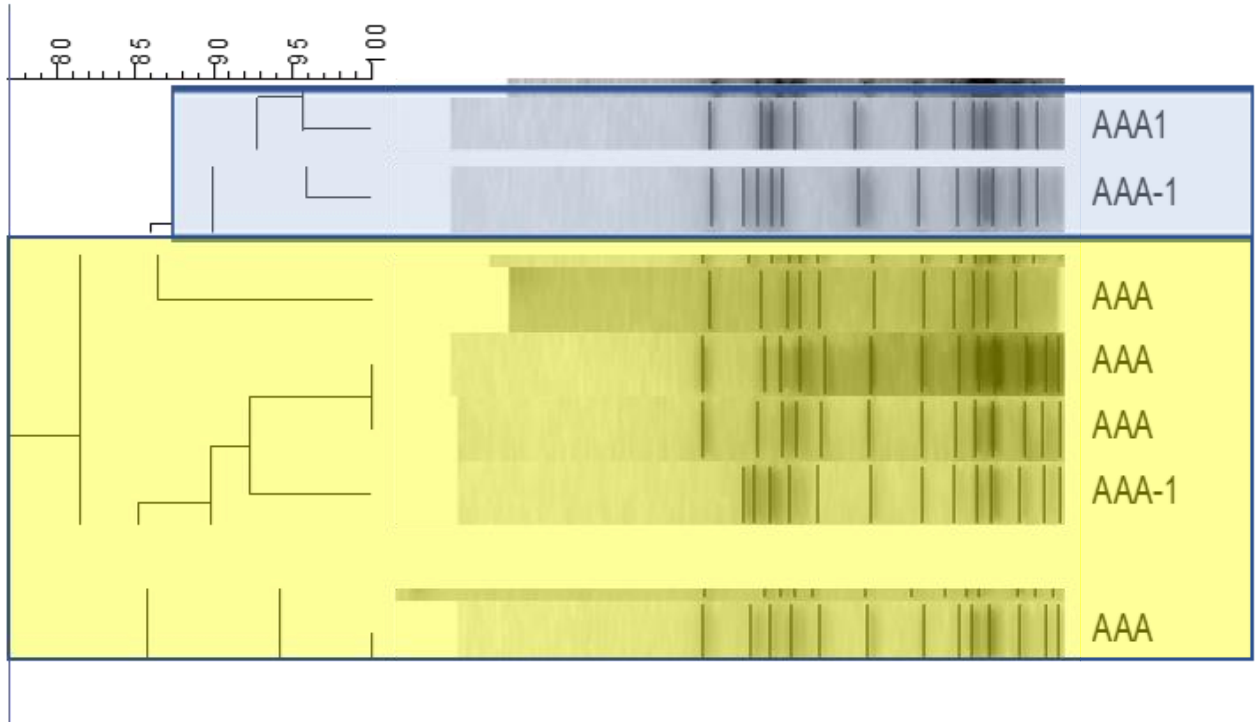


Figure 5.8 AAA pulsotype isolates and their relative similarity based on PFGE, colored based on spectrotyping results

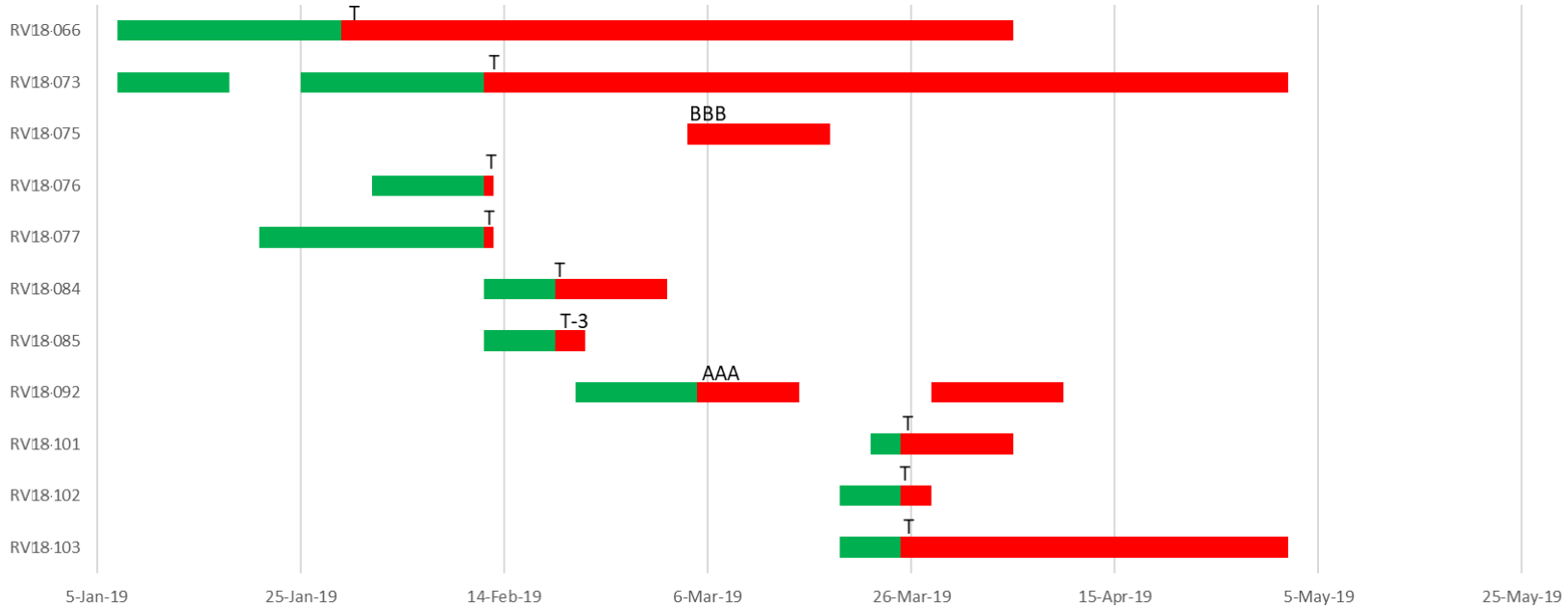


Within 12 wards that were analysed at RVH, at the end of the study period, 15 clusters of isolates based on spectral similarity were identified (Table 5.11). Of these 15 clusters, 7 clusters were 100 % concordant with PFGE results. FTIR spectrotyping identified 4 additional clusters that were not identified by PFGE. These clusters were comprised of 2-4 isolates all with different pulsotypes, which were considered sporadic strains. The categorical agreement in cluster identification for wards in RVH ranged from 55.6 % to 100 % (Table 5.11). There were a few PFGE pulsotypes that were identified across multiple wards, including pulsotypes V-2, T and BBB, that were identified as clusters by FTIR spectrotyping. In wards C09S and D03, multiple clusters were observed, indicating potential presence of two outbreaks that were ongoing at the same time. Environmental isolates were acquired in RVH units, C08S and C09C and C09S. Isolation from unit environment indicates persistent presence of VRE. In the case for C08S, environmental isolates were clustering with the isolates that were identified as an outbreak. Environmental samples were typed during the outbreak in late February 2019. At this point, 5 patients were screened positive during their stay in the ward, where none of these patients were previously colonized with VRE. During the time of the study, the VRE screening also revealed that three more patients that had stayed at the unit in mid to late March also acquired the same pulsotype VRE, indicating that the outbreak was still ongoing (Figure 5.9). This was later confirmed by PFGE, where all isolates that clustered by FTIR were identified as pulsotype T. In Table 5.10, consecutive analysis by date in which new VRE positive patients are identified or is moved into the ward from other wards. The analyses provided concordance to PFGE results at every analysis, with the exception of one, where a sporadic isolate (pulsotype BBB) was considered clonal to the outbreak isolates. However, interestingly, the once considered clonal isolate was determined to be non-related to the outbreak in consecutive analyses.

In ward C09S, environmental sample ES37 clustered with VRE samples of patients that had stayed in the unit, indicating again, that the environmental sample and patient samples were clonal. On the contrary, in ward C09C, spectrotyping results indicated that the environmental sample, ES54 was a unique, and sporadic isolate. This result was discordant to PFGE results, which indicated that the environmental sample was clonal to a sample, that acquired VRE colonization as of February 18th. The environmental sample was collected and analysed on March 5th. This unit resulted in two clusters, one which comprised of a CC and CC-1 isolate (n = 2), and another of 4 sporadic isolates (DDD, HH-3, LL-2 and QQ-1). Two of the four sporadic isolates that clustered

together were from patients who were already VRE colonized prior to staying in the ward, indicating that the transmission did not occur in this ward. However, it does not preclude the idea that the same strain type VRE colonized all four patients, as the same strain type VRE may have adapted and persisted in the environment across wards within the hospital. Based on the FTIR results, two outbreaks were ongoing in this subunit, however, it is likely that only one patient was colonized from a VRE transmitted in this ward.

Figure 5.9 Timeline of patients' stay and VRE colonization and pulsotype at subunit C08S



Example timeline of patient movement within a subunit at a hospital. Green bars represent time during the patients' stay without VRE colonization. Red bars represent time spent in the subunit with VRE positive colonization. PFGE pulsotype indicated above initial colonization time

Table 5.12 Results by date in Subunit C08S over a 4-month period, by FTIR spectrotyping and PFGE

Isolates analysed (N)	Date	Added New Sample ID	PFGE pulsotype designation	Result by FTIR spectrotyping	Outbreak Detection by FTIR spectrotyping (Y/N)	Result in comparison to PFGE
4	12-Feb-19	RV18066, RV18073, RV18076, RV18077	T x4	clonal	Yes	concordant
6	19-Feb-19	RV18084, RV18085	T x 5 T-3	Clonal for all T samples. New sample (#085, T-3) not clonal	Yes	concordant
8	27-Feb-19	ES11, ES13	T x 7/T-3	clonal for all T samples.	Yes	concordant
9	04-Mar-19	RV18075	T x 7/ T-3/BBB	new sample clonal to all T samples	Yes	not concordant
10	05-Mar-19	RV18092	T x 7/T-3 /BBB/AAA	Sporadic. Previously identified clonal sample was not clonal in this analysis	No new isolates part of outbreak	concordant
13	25-Mar-19	RV18101, RV18102, RV18103	T x 7/T-3 /BBB/AAA/ T x 3	new samples all clonal to T samples		concordant
14	01-May-19	Non-hospital acquired isolate	T x 7/T-3 /BBB/AAA/ T x 3/non-hospital acquired	new sample is sporadic		concordant

5.4.4 Discussion

The FTIR spectrotyping method was developed based on the knowledge of VRE strain types provided by PFGE analysis. Using the PFGE results of retrospective isolates, a cut-off value for spectrally "clonal" isolates was determined (500 squared Mahalanobis distance). The PCA-LDA method does not rely on any prior knowledge on strain type or relatedness between samples. Using the spectral replicates for each isolate that were analysed, distances between isolates were maximized by identification of spectral regions that contribute to differentiation of isolates. This method was developed on the idea that spectral similarities between clonal isolates should be high even with the PCA-LDA model undergoing an optimization in maximizing distances between such isolates, because the clonal isolates should be spectrally indistinguishable. PCA was conducted first prior to LDA for each analysis, to ensure that isolates are not differentiated on variables other than the microbial differences, such as instrumental variability. Additionally with variable reduction, LDA can be completed faster compared to when analysis is conducted using all datapoints that are present in the spectral region 1350-800 cm^{-1} (1181 datapoints). Discrimination in a PCA-LDA between isolates of the two most commonly acquired pulsotypes AA and CC, resulted in clustering of isolates with 91.7 % and 91 % categorical agreement respectively, showcasing the presence of spectral differences amongst VRE isolates with different pulsotypes.

In both the retrospective and prospective studies, FTIR spectrotyping clustered isolates of the same pulsotype with success, reflected by high true positive rates. The false positive rates varied between analyses, indicating that the spectrotyping did not observe sufficient spectral difference between some pulsotypes, while others were easier to discriminate due to larger spectral differences. PFGE patterns are known to be affected in the number of bands present, as well as in the patterns, through the introduction of transposons, and *van* genes into the genome [18]. Since accessory genomes are known to cause large variations in phenotypic characteristics within *E. faecium*, it is no surprise that there are discordant clustering results between FTIR spectrotyping and PFGE. Follow-up analysis that compares FTIR spectrotyping with WGS, as well as correlation between genotypic and spectral characteristics may further aid in understanding the extent in which FTIR spectrotyping can be used and be beneficial as a rapid routine screening method of clonal isolates.

There were a few limitations to this study that must be mentioned. The first limitation in this study was that the VRE samples that were considered non-nosocomial (the patient acquired VRE prior to hospitalization) were not analysed for strain type by PFGE, and these isolates reported as unidentified. Additionally, PFGE analysis was only conducted on the first VRE isolate obtained from screening per patient, or when the patient was VRE positive again after at least 3 months of being VRE-negative. Since PFGE is both labor and time intensive, and impractical to analyze all isolated VRE samples, subsequent VRE samples isolated from patients who were previously identified to be VRE-positive in prior screening were not analyzed by PFGE, despite patients on outbreak wards being screened on a weekly basis (or once every two weeks for patients staying on non-outbreak wards). Thus, these VRE samples isolated in subsequent screenings from VRE colonized patients were also not available for FTIR spectral acquisition. Colonization with multiple *E. faecium* strain types occur commonly [18, 19], and although repeated screenings can identify whether patients are colonized or not, without the strain typing results by PFGE on consecutive samples from the first positive sample, it is not clear whether the patient is still colonized with the same strain type, or has acquired new strain types during hospitalization. Ideally, repeated VRE samples that are isolated from colonized patients should be strain typed and characterized in order to track changes like acquisition of antimicrobial resistance and virulence genes. Although genome diversification is known to occur through optimization to host during colonization and infection, the extent of the microorganisms' genotypic and phenotypic changes are not well known [20]. Frequent strain typing also makes identification of pathogen transmission in the environment and to new patients easier. The advantage of the FTIR spectrotyping method is that it can handle high sample throughput, as no complicated sample preparation are required, and spectral acquisition can be achieved within a few minutes from obtaining isolated colonies on agar plates. Furthermore, the FTIR spectrum reflects the metabolic and structural characteristics of the live, whole-cell of microbial sample, which may detect both genetic and phenotypic changes that occur in microorganisms.

Although common pulsotypes were observed in multiple wards, based on the movements of patients between wards as well as the dates when colonization were first detected for patients, patient to patient transmission was rarely observed. The presence of VRE in the environment, which were spectrally and genotypically clonal to that of patients who stayed in the ward indicate that there are endemic VRE strains. The results show that among the wards in both RVH and MGH,

there were pulsotypes that were present in both hospitals, such as AAA pulsotype, which indicate the potential spread of strain types between hospitals. The FTIR spectrotyping result for AAA pulsotype that was identified in both MGH and RVH during the 6-month prospective study showed that the samples were spectrally indistinguishable, and the clustering was concordant to PFGE results. Multiple endemic VRE strain types were present in both hospitals, which is commonly observed in hospitals [18]. A report on a VRE outbreak at a neonatal intensive care unit in Montreal that occurred in August 2018, indicated that the hand hygiene protocol was only followed by 62 % by hospital workers, patients and their visitors [21] . Education of the patients and workers, as well as visitors was suggested to increase adherence to hygiene protocol and result in better control over the outbreaks. The current cleaning and hand washing policies that are implemented when VRE positive patients are identified at the hospital include 1) daily cleaning of surface areas that are frequently touched, 2) disinfection of all equipment with approved disinfectant, and 3) hand washing must be done with alcohol based rinse, or with soap and water. When isolation precautions are removed, a final cleaning is required for the ward where the patient had been hospitalized. VRE have been reported to be resistant against chemicals used in standard cleaning procedures, which makes decontamination harder, as they continually adapt to and persist in their environment [2, 22].

The FTIR spectrotyping method is indeed useful when an outbreak is occurring, and furthermore, it is useful in identifying a sample with unusual spectral characteristic. However, it should be noted that there are reduced levels of confidence in identifying clusters when more than one outbreak is simultaneously occurring. Although not impossible to identify more than one cluster of isolates in a unit (like Unit 17 and 15 at MGH), it seems that in wards like C09C, C09S and C10 where more than 6 pulsotypes of VRE isolates were present, it was difficult to discriminate between isolates well. However, due to the lack of additional genotypic information on the isolates (such as MLST sequence type and the types of genes each isolate possesses for virulence and pathogenicity) it is difficult to determine whether these various pulsotypes are part of a broader MLST sequence type, or are phylogenetically more similar to each other than to other strain types. The knowledge on *E. faecium* strain types is limited, and expected to expand and change as more isolates are whole genome sequenced and recategorized [20].

When comparing samples, it assumed that the VRE strain type that colonized patients were consistently the same throughout their hospitalization, and that only one strain type colonized the

patient, rather than possibly multiple types. The results for clustering was also based on the assumption that there were no genetic and phenotypic evolutions that occurred while colonizing patients. VRE has a high recombinant rate compared to other organisms, and therefore, this study could be improved through the spectral data acquisition from every positive sample that the screening identified. To address these limitations, future studies should acquire FTIR and genotypic data from patients repeatedly to track the changes, and evaluate whether FTIR spectroscopy is able to rapidly identify outbreaks or odd samples, that can be further analyzed for characterization, to provide a better global understanding from typing results that would better track the transmission and evolution of VRE colonization in patients.

5.5 Conclusion

In this study, transflection FTIR spectroscopy demonstrated its ability to identify clusters of bacterial isolates based on spectral similarity using VRE isolates for outbreak surveillance. Using unsupervised and supervised multivariate statistical analyses, spectral discrimination of two VRE strain types (previously identified by PFGE as AA and CC) was successfully achieved. A unsupervised PCA-LDA method, with a cut-off criteria to determine spectrally indistinguishable isolates was developed using the AA and CC pulsotype VRE isolates, resulting in 91.7 % and 91 % concordant clustering relative to PFGE results. The method was evaluated on 214 retrospective isolates that were collected from patients who acquired VRE from 10 different wards, and in prospective studies with 129 isolates from two Montreal hospitals. Analyses indicated that the FTIR spectrotyping method had high true positive rates relative to PFGE results, showcasing its ability to identify clusters of VRE isolates that were spectrally indistinguishable. The detection of VRE clusters by FTIR spectrotyping which were later confirmed by PFGE in the prospective study demonstrated its utility as a routine method in the clinical setting. The discrepant results in clustering between FTIR spectrotyping and PFGE could not be studied by a third method like WGS, which could provide genotypic insight, such as sequence types by MLST, as well as SNP differences. WGS analyses may provide insight into genotypic characteristics that may be related to spectral differences (or the lack thereof) observed between VRE isolates that had discrepant results between FTIR spectrotyping and PFGE pulsotype analyses. FTIR spectrotyping would serve as a useful tool in rapidly screening and identifying clusters of isolates that may be related to an outbreak, as well as spectrally unique strains, which can then be further characterized by genotypic methods such as whole genome sequencing.

5.6 References

1. CDC, *Antibiotic resistance threats in the United States, 2019*. 2019, U.S. Department of Health and Human Services: Atlanta, GA.
2. Zhou, X., et al., *Enterococcus faecium: from microbiological insights to practical recommendations for infection control and diagnostics.* Antimicrobial Resistance & Infection Control, 2020. **9**(1): p. 1-13.
3. Salipante, S.J., et al., *Application of whole-genome sequencing for bacterial strain typing in molecular epidemiology.* Journal of clinical microbiology, 2015. **53**(4): p. 1072-1079.
4. Novais, Â., et al., *Fourier transform infrared spectroscopy: unlocking fundamentals and prospects for bacterial strain typing.* European Journal of Clinical Microbiology & Infectious Diseases, 2019. **38**(3): p. 427-448.
5. Gilmore, M.S., et al., *Enterococci: From Commensals to Leading Causes of Drug Resistant Infection [Internet]*. 2014.
6. Schürch, A., et al., *Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene–based approaches.* Clinical Microbiology and Infection, 2018. **24**(4): p. 350-354.
7. Vogt, S., et al., *Fourier-transform infrared (FTIR) spectroscopy for typing of clinical Enterobacter cloacae complex isolates.* Frontiers in microbiology, 2019. **10**.
8. Sousa, C., et al., *Development of a FTIR-ATR based model for typing clinically relevant Acinetobacter baumannii clones belonging to ST98, ST103, ST208 and ST218.* Journal of Photochemistry and Photobiology B: Biology, 2014. **133**: p. 108-114.
9. Rodrigues, C., et al., *On the front line of Klebsiella pneumoniae surface structures understanding: establishment of Fourier Transform Infrared (FT-IR) spectroscopy as a capsule typing method.* bioRxiv, 2019: p. 554816.
10. Dinkelacker, A.G., et al., *Typing and Species Identification of Clinical Klebsiella Isolates by Fourier Transform Infrared Spectroscopy and Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry.* Journal of clinical microbiology, 2018. **56**(11): p. e00843-18.
11. Dieckmann, R., et al., *Rapid characterisation of Klebsiella oxytoca isolates from contaminated liquid hand soap using mass spectrometry, FTIR and Raman spectroscopy.* Faraday discussions, 2016. **187**: p. 353-375.
12. Grunert, T., et al., *Rapid and reliable identification of Staphylococcus aureus capsular serotypes by means of artificial neural network-assisted Fourier transform infrared spectroscopy.* Journal of clinical microbiology, 2013. **51**(7): p. 2261-2266.
13. Silva, L., et al., *Fourier transform infrared (FT-IR) spectroscopy typing: A real-time analysis of an outbreak by carbapenem-resistant Klebsiella pneumoniae.* European Journal of Clinical Microbiology & Infectious Diseases, 2020. **39**(12): p. 2471-2475.
14. Rakovitsky, N., et al., *Fourier Transform Infrared Spectroscopy is a new option for outbreak investigation: A retrospective analysis of an Extended-Spectrum-Beta-Lactamase-producing Klebsiella pneumoniae outbreak in a neonatal intensive care unit.* Journal of clinical microbiology, 2020. **58**(5): p. e00098-20.
15. AlMasoud, N., et al., *Rapid discrimination of Enterococcus faecium strains using phenotypic analytical techniques.* Analytical Methods, 2016. **8**(42): p. 7603-7613.

16. Morrison, D., et al., *DNA banding pattern polymorphism in vancomycin-resistant Enterococcus faecium and criteria for defining strains.* Journal of clinical microbiology, 1999. **37**(4): p. 1084-1091.
17. Johler, S., et al., *High-resolution subtyping of Staphylococcus aureus strains by means of Fourier-transform infrared spectroscopy.* Systematic and applied microbiology, 2016. **39**(3): p. 189-194.
18. Bonten, M.J., et al., *Stability of vancomycin-resistant enterococcal genotypes isolated from long-term-colonized patients.* Journal of Infectious Diseases, 1998. **177**(2): p. 378-382.
19. Tremlett, C.H., D.F. Brown, and N. Woodford, *Variation in structure and location of VanA glycopeptide resistance elements among enterococci from a single patient.* Journal of Clinical Microbiology, 1999. **37**(3): p. 818-820.
20. Freitas, A.R., et al., *Multidrug-resistant high-risk Enterococcus faecium clones: can we really define them?* International Journal of Antimicrobial Agents, 2020: p. 106227.
21. Stachura, M. and M.-A. Lefebvre, *Epidemiologic Investigation of an Outbreak of Vancomycin-resistant Enterococcus in a Tertiary Care Neonatal Intensive Care Unit.* American Journal of Infection Control, 2020. **48**(8): p. S37.
22. Ramos, S., et al., *Enterococci, from harmless bacteria to a pathogen.* Microorganisms, 2020. **8**(8): p. 1118.

Connecting Statement

Successful discrimination between pulsotypes of VRE isolates related to nosocomial outbreaks was demonstrated, using transflection-FTIR spectroscopy. A method was developed to identify isolates that were considered clonal based on spectral similarity. The method was tested on prospective samples and like the retrospective isolates, proved the method to successfully identify clusters of isolates that are spectrally clonal. The method's ease of sample preparation, data acquisition, and analysis demonstrated its potential use as routine typing technique, possibly in replacement of PFGE, and as a screening method prior to genotypic methods like WGS. However, owing to the extensive overlap in FTIR spectral regions where absorbances are observed for different biochemical constituents from derived from structural and metabolic features of the microorganisms, it is near impossible to identify the exact biomolecules that contribute in successful spectral discrimination. In the following chapter, high-resolution magic-angle spinning (HR-MAS) NMR spectroscopy, a spectroscopic technique with higher chemical specificity was used to acquire and spectrally differentiate microorganisms based on their highly resolved NMR spectra from whole, live cells. Utilizing this technique in combination with whole genome sequencing and mass spectrometry, a tentative assignment of biomolecules that enabled spectral differentiation between *E. faecium* strain types by FTIR spectroscopy were identified.

Chapter 6. Preliminary data acquired for correlation between genotypic and spectroscopic data for biomarker elucidation and correlation to FTIR spectral analysis, using NMR spectroscopy and whole genome sequencing

6.1 Abstract

High-resolution magic-angle spinning nuclear magnetic resonance (HR-MAS NMR) spectroscopy is a whole-organism fingerprinting technique similarly to FTIR spectroscopy, capable of acquiring spectral data from live microbial cells. While both techniques analyze the overall biochemical composition of live cells, NMR spectroscopy has gained popularity in studying metabolomics and cell-wall structures of intact microorganisms owing to its high chemical specificity and rich information content. In the present study, ^{1}H and ^{31}P HR-MAS NMR spectroscopy were explored for their discriminatory capabilities, for potential in rapid identification and strain typing. Genus and species discrimination were demonstrated using 46 reference and clinical strains of yeasts and Gram-positive and Gram-negative bacteria. Furthermore, 28 vancomycin resistant *Enterococcus faecium* (VRE) isolates were selected to examine the potential applicability of ^{1}H and ^{31}P HR-MAS NMR spectroscopy for discrimination amongst isolates from three nosocomial outbreaks. Following overnight culture on blood agar, sufficient microbial mass was collected and placed in an 80µl HR-MAS NMR insert which was then put in a 4mm MAS rotor. Spectral data analyses were performed by hierarchical cluster analysis (HCA) and principal component analysis (PCA) in conjunction with the use of feature selection algorithms. Classification of the VRE isolates by both PCA and HCA of the ^{1}H or ^{31}P HR-MAS NMR spectra yielded 100 % and 96 % concordance with PFGE results respectively. This proof-of-concept study demonstrated how genotypic differences among strain types are reflected in phenotypic differences that can be observed by ^{1}H and ^{31}P HR-MAS NMR whole-organism fingerprinting techniques and supports the potential development of HR-MAS NMR spectroscopy as a strain typing tool for nosocomial outbreak surveillance. Additionally, owing to the increased chemical specificity of HR-MAS NMR relative to FTIR spectroscopy, tentative biomarker identification was achieved using ^{1}H and ^{31}P NMR spectroscopy, in combination with genotypic data acquired from whole genome sequencing (WGS) and mass spectrometric (MS) data, that differentiates between the VRE strain types

associated with three different nosocomial outbreaks. Phosphorus and choline containing molecules were identified as potential metabolic or structural differences that were spectrally observed between the VRE pulsotypes, in FTIR and NMR spectral analyses. Additionally, WGS data revealed that the genotypic differences between the strain types were extensive, with the pulsotypes only sharing 27 % of the core genome. CC pulsotype isolates were missing a housekeeping gene, *pstS*, which encodes for a protein that binds to phosphate for uptake. Extended studies in the future are required to confirm the identity of tentatively identified biomarkers with additional multidimensional NMR experiments and targeted mass spectrometry. This work provided a list of potential biomarkers that result in differences observed in FTIR spectral characteristics between VRE isolates at the subspecies level, making a strong case for rapid and routine strain typing by FTIR spectroscopy as a screening method.

6.2 Introduction

As presented in the previous chapters and by many other researchers, microbial discrimination and identification at species and sub-species level can be achieved using FTIR spectroscopy in combination with multivariate statistical analyses [1-13]. Due to overlapping absorption signals from various biochemical constituents of the microorganisms' structural and metabolic composition, specific biomarkers or biochemical differences that enable spectral discrimination are seldom identified. In addition to reporting the capabilities of FTIR spectroscopy for rapid microbial identification and analyses, to become a widely accepted method, it is important to explain the observed infrared spectral differences by supplementing the data with other techniques such as nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS).

6.2.1 NMR Spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is a well-established analytical technique used in many different fields including chemistry, particularly organic chemistry, physics and biochemistry. The most commonly used nuclei include ^{1}H, ^{13}C, ^{15}N and ^{31}P when studying biological samples [14]. Common uses include elucidation of molecular structures, amino acid sequence in proteins, and studies of reaction rates and dynamics. NMR spectroscopy is also applied in studying metabolomics, observing the state of and the change in metabolism under different conditions. Acquiring high resolution NMR spectra in solution-state and solid-state NMR spectroscopy, traditionally required that samples be in a homogeneous liquid form and crystalline solid state respectively, limiting the types of samples and the applications for which NMR spectroscopy was suited. The development of the technique known as high-resolution magic-angle spinning (HR-MAS) NMR spectroscopy provided the ability to analyze much more diverse sample types, including semi-solids, and heterogenous samples, such as microorganisms, tissues, and whole cells, as well as food and beverage products [15-20]. This increased NMR spectroscopy's popularity in metabolomic studies, expanding into disciplines such as food science, agricultural science, and medical studies. HR-MAS NMR spectroscopy has the ability to acquire high-resolution spectra that look similar to spectra acquired in solution-state NMR spectroscopy through the suppression of line broadening caused by dipole-dipole interactions and chemical shift anisotropy. This is achieved by spinning the sample at the "magic angle" relative to the external

magnetic field, where the magic angle $\theta_m = \sim 54.74°$, was determined from the following equation: $\cos^2\theta_m = \frac{1}{3}$ [21]. HR-MAS NMR spectroscopy allows samples like bacteria to be analyzed without preparation steps such as extraction, purification and concentration which were required for solution-state and solid-state NMR experiments. Not only does this shorten the sample preparation time substantially, but it also allows analysis of the sample in its most natural state. The advantage of NMR spectroscopy over vibrational spectroscopic techniques (FTIR and Raman) is that it has superior chemical specificity, allowing the presence and identity of metabolites to be determined from the NMR spectra, especially through the use of multidimensional NMR experiments.

6.2.1.1 Applications of NMR spectroscopy in microbial identification.

Successful microbial discrimination at genus and species level has been demonstrated in solution-state, where microbial colonies were suspended in solution, or the spectra were acquired from the broth media after growth and removal of microbial colonies [22-25]. In order to acquire high resolution spectra in the solution-state, bacteria and yeasts were suspended in liquid broth, and spectra were acquired before the microorganisms sedimented to the bottom of the NMR tube. The experiments conducted on the broth media relied on observations of the metabolites that were released by the bacterial species, in order to discriminate between organisms. These studies showcase the potential applicability of HR-MAS NMR spectroscopy in rapid identification and strain typing of microorganisms, directly from live cells. At the time of writing, there have yet to be any studies conducted in ^{1}H or ^{31}P HR-MAS NMR spectroscopy, for microbial identification or strain typing. However, multidimensional HR-MAS NMR experiments have been popular in studying microorganisms. The potential in serotyping was reported through identification of polysaccharide moieties on cell wall structures of live *Neisseria meningitidis* and *Candida albicans*, glycans on *Campylobacter* species and O-specific polysaccharides on *Yokenella regensburgei*, which were matched with *in vitro* analyses [26-29]. By acquiring the data from microorganisms in their live state, HR-MAS NMR spectroscopy can be used to study the interaction between the microorganism and desired factors, such as the effect of drugs, gene mutation, osmotic pressure, aging and nutrient variation [15, 28, 30-33].

6.2.2 Mass spectrometry for metabolomics

Mass spectrometry is a complementary technique to NMR spectroscopy in metabolomics, as it has better sensitivity, with a lower limit of detection, ranging from pM to nM [34]. While NMR spectroscopy can identify metabolites with little or no prior knowledge, MS can identify metabolites through the measured mass, to determine the elemental composition and molecular formula [34].

6.2.3 Genotypic techniques for strain typing and characterization

Genotypic techniques used for VRE strain typing includes methods like pulsed field gel electrophoresis (PFGE), multiple locus variable number of tandem repeat analysis (MLVA), multi-locus sequence typing (MLST), whole genome sequencing (WGS) and transposon analysis [35]. Core genome MLST (extracted from WGS data) has excellent discriminatory power, and reproducibility, as well as improved ease of performance and data exchange compared to PFGE. With the increasing popularity of WGS, cost is expected to decrease, making it a viable option upon standardization and validation of the WGS procedure and data interpretation. WGS is also beneficial due to its wealth of information, such as the possession of virulence factors, resistance genes, plasmids and other characteristic genetic markers that may be of interest [35]. Especially in the case of *E. faecium*, which has high genetic plasticity, frequent acquisition and deletion of genes through plasmid transfers can be observed through analysis of the whole genome data.

6.3 Materials and Methods

6.3.1 Clinical isolates used for evaluating the discriminatory capabilities of ^{1}H spectra acquired by HR-MAS NMR spectroscopy

Forty-six isolates, including 12 ATCC strains were chosen to demonstrate the discriminatory capabilities of ^{1}H HR-MAS NMR spectroscopy at the Gram stain, genus and species levels, as well as sub-species level, demonstrated by discrimination of MRSA and MSSA. All isolates were cultured from frozen 10 % glycerol stock onto Columbia agar with 5 % sheep blood (BAP) (Oxoid, Nepean, ON) and incubated for 24 hours at 35˚C. Prior to spectral acquisition, isolates were sub-cultured onto BAP and grown under the same conditions.

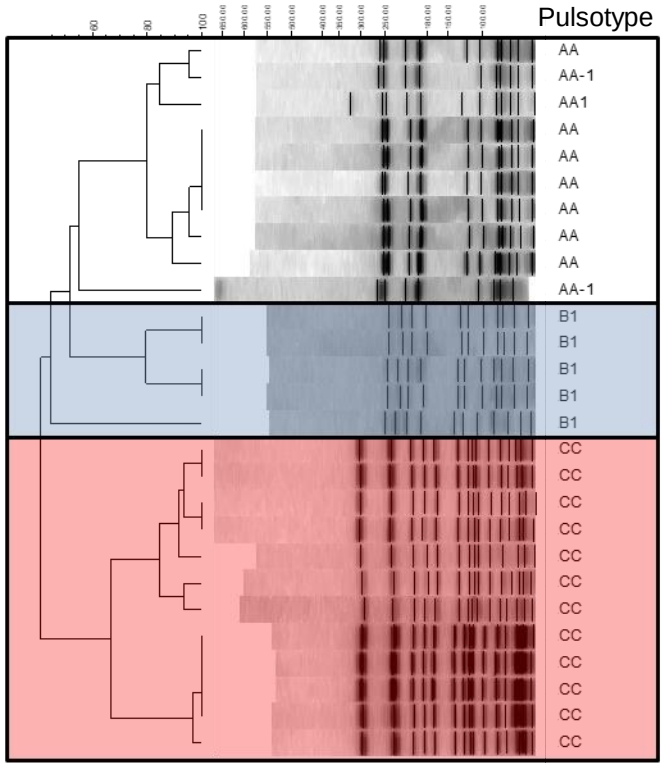
6.3.2 VRE isolates used for evaluating the discriminatory capabilities of ^{1}H and ^{31}P NMR spectra acquired by HR-MAS NMR spectroscopy and tentative biomarker identification relating to strain-type differences in spectral data

Vancomycin-resistant *E. faecium* (VRE) isolates (*n* = 28) originating from 3 different nosocomial outbreaks at the Montreal General Hospital (MGH) between April 2016 to December 2017 were used to evaluate the potential applicability of ^{1}H and ^{31}P HR-MAS NMR spectroscopy as rapid strain typing techniques retrospectively. All isolates were previously identified by routine clinical laboratory identification methods (VRE screening broth, PCR and chromogenic VRE agar), and analyzed by PFGE for their strain type characterization (Figure 6.1and Table 6.1). Isolates were prepared from frozen 10 % glycerol stock, by culturing on Columbia agar with 5 % sheep blood (BAP) (Oxoid, Nepean, ON), and incubating at 35°C for 18-24 hours. They were sub-cultured on BAP under the same conditions prior to spectral acquisition by FTIR and HR-MAS NMR spectroscopy. Cultures on VRE ChromAgar plates (Biomeriux, FR) were checked to confirm the isolates' vancomycin resistance, by the color of the grown colonies (blue/purple for *E. faecium*).

Table 6.1 PFGE pulsotypes of vancomycin resistant E. faecium isolates used in study for discrimination by HR-MAS NMR spectroscopy.

PFGE Pulsotype	**Isolates**	**Outbreak Dates**
AA	11	ICU/Unit 12E Sep-Oct 2016
B1	5	Unit 11E Jan 2017
CC	12	Unit 17E/15 Jan – March 2017

Figure 6.1 Dendrogram generated from hierarchical cluster analysis using PFGE band patterns of VRE isolates from three hospital outbreaks



HCA of PFGE band patterns identified three clusters (AA, B1 and CC), that occurred from September 2016 to March 2017. Axis on the dendrogram arm indicates percent similarity between isolates' PFGE band patterns. AA showed highest degree of variability amongst the isolates within each pulsotype, noted by (-) 1 in the PFGE pulsotype nomenclature when an extra band is present or missing.

6.3.3 ^{1}H and ^{31}P HR-MAS NMR spectral acquisition and processing parameters

The ^{1}H HR-MAS NMR spectra of bacteria were acquired using an AVANCE III 600 MHz NMR spectrometer (Bruker Corporation, GE) equipped with a room temperature (25 ℃) HCP and HCN Z-gradient HR-MAS probes. All spectra were acquired by co-adding 256 scans, performed with pre-saturation of the water peak and a pulse at 90º (pulse program zgpr). Disposable 80µl inserts were filled half-way with bacterial colonies directly from agar, using a centrifuge at 10000 rpm. 3-(Trimethylsilyl)propanoic acid (TSP) was dissolved in D_2O at a concentration of 1 % (w/v) and used as an internal standard. A 5 µl aliquot of the TSP solution was pipetted into the rotor

before the bacteria-filled insert was put in. This prevented any potential H-D exchange between the D_2O and the bacteria. Due to the lack of D_2O in the rotor relative to the sample, the field was locked on a separate insert containing 1:1 ratio of H_2O/D_2O solution. The sample was spun at 5 kHz at the magic angle (54.7°) relative to the external magnetic field. Each spectrum was phased, referenced (TSP at δ = 0 ppm), and baseline-corrected using the Topspin Software ver. 3.6.3 (Bruker Corporation, Billerica, MA, USA).

Phosphorous HR-MAS NMR spectra were spectra were acquired with 512 coadded scans using the pulse program zg, spinning at 5 kHz at the magic angle (~54.74°) relative to the external magnetic field. A rotor containing triphenyl phosphate (TPP) dissolved in acetone was used as an external standard at -27.5 ppm. A ^{31}P HR-MAS NMR spectrum of the external standard was acquired once a day, before or after sample spectral acquisitions. The spectra of samples were referenced according to the spectral shift observed in the TPP spectrum acquired on the same day.

The post-processed NMR data were subjected to statistical analysis using AMIX software (Bruker Corporation, Billerica, MA, USA). PCA was employed on the NMR spectra using the region δ = 0.5 ppm to δ = 10ppm, excluding the water peak region (δ = 4.5-5.2 ppm). Additionally, the NMR spectra were pre-processed using vector normalization and hierarchical cluster analysis (HCA) was employed to visualize the spectral similarities between NMR spectra of different microbial organisms, using SpectrAnalysis software (Cognisolve Inc., Montreal, QC).

6.3.4 Transflection FTIR spectral acquisition and data processing parameters

Is Isolated colonies were picked directly from agar plates using sterile 1-µl disposable loops and smeared onto IR-reflective, low-E slides (Kevley Technologies, Chesterland, OH, USA) as thin films in triplicate spots and air-dried. Using the Cary 630 FTIR spectrometer (Agilent Technologies, Santa Clara, CA) with a 10° specular reflectance sample interface (Agilent Technologies, Santa Clara, CA) or a SurveyIR microscope (Czitek, Danbury, CT, USA) and MicroLab Software (Agilent Technologies, Santa Clara, CA), trasnflection spectra were acquired with 64 coadded scans at 8 cm^{-1} resolution and a zero-filling factor of 8, in the spectral range 4000-650 cm^{-1}. Prior to spectral acquisition, background spectra were acquired from a bare surface area of the E-glass slide with the same spectral acquisition parameter settings. Spectra were pre-processed by taking the first derivative and vector normalization to eliminate variability caused by baseline shifting and sample thickness respectively.

Using the spectral region 1350-800 cm^{-1}, the spectra were subjected to hierarchical cluster analysis (HCA) and principal component analysis (PCA) to ascertain whether there were any global differences between the two strain types, using JMP Pro ver. 15.2.0 (Cary, North Carolina, USA). PCA was also used to reduce the dimensionality of the spectral data for subsequent multivariate analyses. Analysis to determine spectral relatedness between isolates was conducted with modifications to that reported by AlMasoud [36]. The scores of every PC that explained over 1 % of total variance were used for subsequence linear discriminant analysis (LDA). Unsupervised PC-LDA developed a model that maximized the distance between isolates (where each isolate was classified as its own group), and minimized the distance between replicate spectra per isolate. Squared Mahalanobis distances for each spectrum to the mean of each class, were calculated from the PC-LDA model. Using the obtained squared distances, isolates were clustered by HCA with average linkage and the results were plotted in a dendrogram to visualize the spectral similarity between isolates. Additionally, using a forward-search spectral feature selection algorithm in SpectrAnalysis software (Cognisolve Inc., Montreal, QC), spectral regions within the fingerprint region (1350-800 cm^{-1}) that contribute to spectral differentiation between the three pulsotypes were identified in intervals of minimum 10 cm^{-1}.

6.3.5 Phenotypic antimicrobial susceptibility test

Using VITEK 2 AST, all VRE isolates were tested to obtain the antimicrobial resistance profile for select antibiotics. The tested antibiotics were benzylpenicillin, gentamycin (high level), streptomycin (high level), quinupristin/dalfopristin, linezolid, tigecycline, ampicillin, ciprofloxacin, levofloxacin, vancomycin, tetracycline, and nitrofurantoin. The resistance status towards each antibiotics was reported in qualitative terms (R = resistant, I = intermediate and S = sensitive), along with quantitative values reported from the determined minimum inhibitory concentration (MIC) in µg/ml.

6.3.6 Genotypic data acquisition using whole genome sequencing of VRE isolates representing AA, B1 and CC pulsotypes

6.3.6.1 Sample Preparation: DNA extraction

Using QIAamp® DNA Mini Kit (Qiagen, Germany), DNA was extracted manually from VRE isolates, as per manufacturer's handbook, based on protocols for bacteria. Briefly, isolates were sub-cultured in Brain Heart Infusion (BHI) broth overnight from the first passage plate

prepared on BAP, from frozen 10 % glycerol stock solution. Bacterial colonies were pelleted using a centrifuge at 5000 x *g* (7500 rpm) for 10 minutes, from 1ml of bacteria broth culture. The bacteria pellet was then suspended in 180µl of enzyme solution, containing 20 mg/ml lysozyme, or 200 µg/ml lysostaphin, 20 mM Tris-HCl (pH 8.0), 2 mM EDTA and 1.2 % Triton, and incubated for 30 minutes at 37 °C. Following the addition of 20 µl of proteinase K and 200 µl of Buffer AL, the sample was vortexed and further incubated at 56°C for 30 minutes, followed by 15 minutes at 95 °C. After 200 µl of 96-100 % ethanol was added, the sample was vortexed prior to applying the mixture to the QIAamp Mini spin column. The sample was then centrifuged at 6000 x *g* (8000 rpm) for 1 minute. The filtrate was discarded, and 500 µl of Buffer AW1 was added to the spin column held in a new collection tube. The sample was centrifuged again at 6000 x *g* (8000 rpm) for 1 minute, and the filtrate was discarded. Buffer AW2 (500 µl) was added to the spin column placed in a new collection tube and centrifuged at 20,000 x *g* (14,000 rpm) for 3 minutes, and the filtrate was discarded. The spin column containing the sample DNA was then transferred to a clean microcentrifuge tube, and 200 µl of Buffer AE was added, and incubated at room temperature for 1 minute, prior to centrifugation at 6000 x *g* (8000 rpm) for 1 minute, to obtain the DNA fraction.

6.3.6.2 Whole genome sequencing

The extracted DNA fragments were quantified using the Quant-iT^{TM} High Sensitivity Assay Kit (Life Technologies Inc., Burlington, ON, Canada) according to manufacturers' recommendations. Sequencing libraries were constructed using the Nextera XT DNA sample preparation and Nextera XT Index Kits (Illumina Inc., San Diego, CA, USA). Paired-end sequencing was performed on the Illumina MiSeq platform, using 600-cycle MiSeq reagent kits (v3) with 5 % PhiX Control (Illumina Inc., San Diego, CA, USA)

6.3.6.3 Quality control and data analysis of genomic data

The acquired raw sequencing data was evaluated on the quality, using FastQC version 0.11.8. Sequences were checked for contamination using ConFindr version 0.5.0. Identification of the microorganism was confirmed based on the sequence, by identifying to the closest match in MASH Reference Genome. Based on quality control parameters (N50 value, number of contigs, total genome length and average coverage depth), each sequencing was given a pass or fail. Single nucleotide variant (SNV) analysis was conducted with all 25 isolates, with sample SEQ164 as the reference isolate. Three additional SNV analyses, one for each cluster (0-1 SNV between samples)

observed in the first SNV analysis was conducted, with 4, 11 and 7 isolates respectively. Antimicrobial resistance genes were identified using ResFindr v4.1 [37]. VirulenceFinder [38, 39] and PlasmidFinder [40, 41] were also used to identify genes relating to virulence and plasmids respectively. Lastly, using the Web database for multi-locus sequence typing from the Center for Genomic Epidemiology, MLST sequence types were determined for the isolates [42].

6.3.7 Preliminary data acquisition of fractionated components of VRE isolates of AA and CC pulsotype

6.3.7.1 Cell lysis and fractionation

Isolated colonies were harvested from BAP using L-shape spreaders, into sterilized glass tubes. Solvents were added to the harvested colonies. First, 1 ml of methanol-d_4 (Sigma-Aldrich, Missouri, USA) and 1 ml of chloroform-d (Sigma-Aldrich, Missouri, USA) were added, and the sample mixture was vortexed. Samples were then ultrasonicated for 15 minutes using Vibra Cell Ultrasonic Liquid Processors Model VCX 130PB 130 Watt (Sonics & Materials Newtown, CT, USA), with settings: 20 % amplitude at room temperature. One ml of D_2O was added and the mixture was vortexed and ultrasonicated with the same setting for 30 minutes. Sample mixtures were transferred to sterilized microcentrifuge tubes, and centrifuged for 10 minutes at 10,000 x *g*. The separated solvent phases were then pipetted out into separate tubes, and stored at -20°C until ready for analysis.

6.3.7.2 FTIR spectral acquisition

FTIR spectra of extracted $CDCl_3$ and D_2O/methanol-D phases were deposited in 1µl quantities on IR reflective, low-E glass slides (Kevley Technologies, Chesterland, OH, USA). Solvents were evaporated off, and deposits were layered until peaks of sufficient absorbance (ie. ~0.30 absorbance units in one or more spectral region of the spectrum) were observed in real-time. Spectra were acquired with 64 coadded scans, 8 cm^{-1} resolution and a zero-fill factor of 8 in the region 4000 to 650 cm^{-1}. Spectral processing and analyses were conducted using OMNIC Spectroscopy Software ver. 9.9.509 (Thermo Fischer Scientific Inc. Waltham, MA, USA).

6.3.7.3 Solution state ^{1}H and ^{31}P NMR spectral acquisition and data processing

An AVANCE III 400 MHz NMR spectrometer (Bruker Corporation, GE) equipped with a room temperature (25 ℃) HCP probe was used to acquire ^{1}H and ^{31}P spectra from extracted

fractions. Samples were transferred into 5mm NMR tubes, with 5 µl of 1 % TSP dissolved in D_2O for water phase samples. Proton NMR spectra were acquired with water suppression and a 90° pulse (zgpr) with 64 transient scans. Phosphorus NMR spectra were acquired with proton-decoupling (zg) and 2048 transient scans. The ^{31}P NMR spectra were referenced to an external reference (TPP dissolved in acetone- d_6) which was acquired on the same day as the spectral acquisition of the samples. Proton NMR spectra were referenced using the deuterated methanol peaks at 3.35 ppm and 4.78 ppm. Spectra were then corrected for baseline shift and phase.

6.3.7.4 MALDI imaging mass spectrometry

MALDI matrices 1,5-diaminonaphtalene (DAN), 2,5-dihydroxybenzoic acid (DHB) and trifluoroacetic acid (TFA) were purchased from Sigma-Aldrich (St-Louis, MO, USA). Liquid chromatography grade solvents were purchased from VWR (Radnor, PA, USA). For statistical analysis of the organic phase samples, one microliter of the sample (in chloroform) was deposited and allowed to dry on a MALDI target. The manipulation was repeated 2 times to concentrate the compounds before matrix deposition. The MALDI matrix DAN was deposited onto the dry droplets using the HTX M3 TM-sprayer connected to an isocratic LC pump using the following parameters: 1,5-DAN solution 5 mg/ml in 67 % acetonitrile; nozzle temperature 65 °C; nozzle height 40 mm; nitrogen pressure 10 psi; flow rate 75 µl/min; z-arm velocity 1200 mm/min; moving pattern VV; track spacing 3 mm; number of passes 26; and drying time 0 s. For the profiling on the aqueous phases (water/methanol), the dried-droplet technique was used. The sample and the matrix were mixed in a 1:1 ratio, and then deposited on the MALDI target. Two MALDI matrices were tested: DAN (10 mg/mL, 67 % acetonitrile) and DHB (20 mg/mL, 50 % methanol). Imaging and profiling were performed on a MALDI TOF/TOF Ultraflextreme mass spectrometer equipped with a SmartBeam II Nd:YAG 355 nm laser operating at 2000 Hz, using the small laser focus setting (Bruker Daltonics, Billerica, MA, USA). IMS data were acquired in positive and negative modes. For imaging MS, 500 shots per pixel were used with a spatial resolution of 200 µm (random walk: 100 shots in a 200 µm diameter) in a mass range of 0-1500 Da. External calibration was carried out using a lipid homemade mix. Post-IMS internal calibration was also performed using matrix signals and known signals. The MALDI IMS data were visualized using the FlexImaging 3.0 software (Bruker Daltonics) and the SCiLS software (2019b Premium 3D, Bruker Daltonics). Lipids were not identified by MS/MS. Only preliminary identification was made through LIPID MAPS based on the exact mass after internal calibration.

6.4 Results

6.4.1 Microbial discrimination using HR-MAS NMR spectroscopy

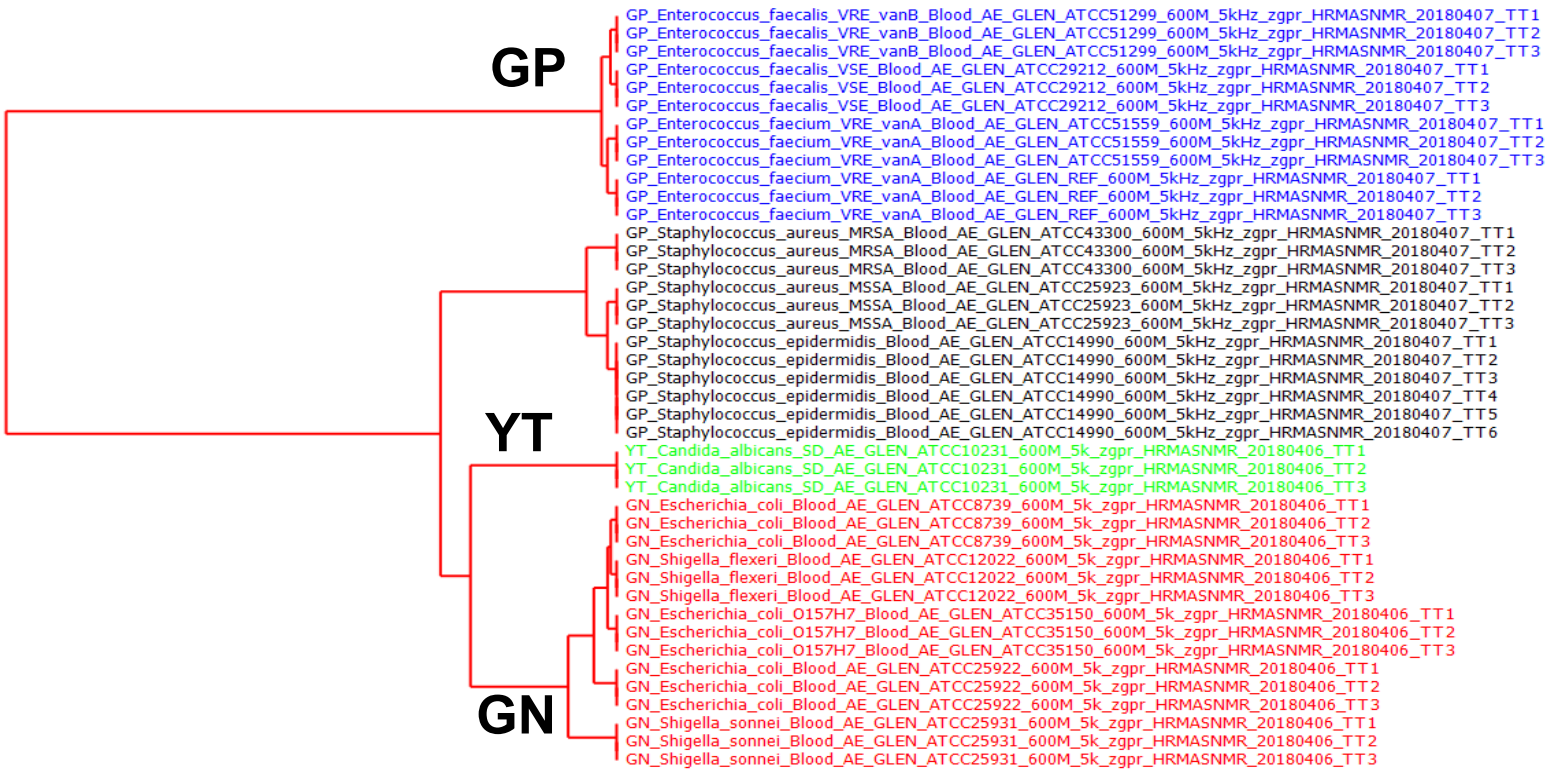
Isolated bacterial colonies that were cultured onto blood agar and MacConkey agar plates were directly filled into HR-MAS NMR inserts, without further sample preparation steps prior to spectral acquisition by ^{1}H HR-MAS NMR spectroscopy. Upon spectral acquisition spectra were referenced using an internal reference (TSP dissolved in D_2O) and corrected for baseline and phase shifts. Successful discrimination between Gram-positive and Gram-negative bacteria was demonstrated using broad chemical shift region (δ = 0.5-4.5 ppm), as visualized in a dendrogram generated from HCA using Ward linkage in Figure 6.2. Furthermore, without spectral feature selection, sufficient spectral differences were observed to discriminate at the genus level (*Staphylococcus, Enterococcus, Shigella, Escherichia, Candida*) and the species level using the broad chemical shift region (Figure 6.2 and Figure 6.3). The sample size in this proof-of-concept study was small, with only 12 ATCC isolates, and therefore extended studies with more isolates per genus and species, as well as a diverse list of genera and species should be conducted to determine the extent of discrimination that HR-MAS NMR spectroscopy is capable of. Spectral discrimination between *Shigella* species and *E. coli* was demonstrated using both blood agar and MacConkey agar grown isolates (Figure 6.4). Like the FTIR spectra, NMR spectra demonstrated that there are biochemical differences between *Shigella* species and *E. coli* that can be observed by spectroscopic techniques in contrast to difficulties encountered by MALDI-TOF MS in the ribosomal protein regions used for microbial identification [12, 43]. The overall spectral discrimination results obtained using the 12 ATCC microbial isolates were in agreement with previous NMR spectroscopy studies that differentiated among various bacteria or yeast samples [22-25, 44, 45]. Differentiation between 9 MSSA and 5 MRSA isolates was also achieved in this study using ^{1}H NMR spectral regions identified by feature selection algorithm within the δ = 0.5-4.5 ppm spectral region, as visualized in the dendrogram (Figure 6.5) generated by HCA, demonstrating subspecies level discrimination.

6.4.1.1 VRE strain typing by ^{1}H HR-MAS NMR spectroscopy

Twenty-eight VRE isolates, originating from 3 different nosocomial outbreaks were utilized to determine if ^{1}H HR-MAS NMR spectroscopy has the discriminatory capabilities for strain typing. These isolates were analyzed and classified as three different strain types using PFGE,

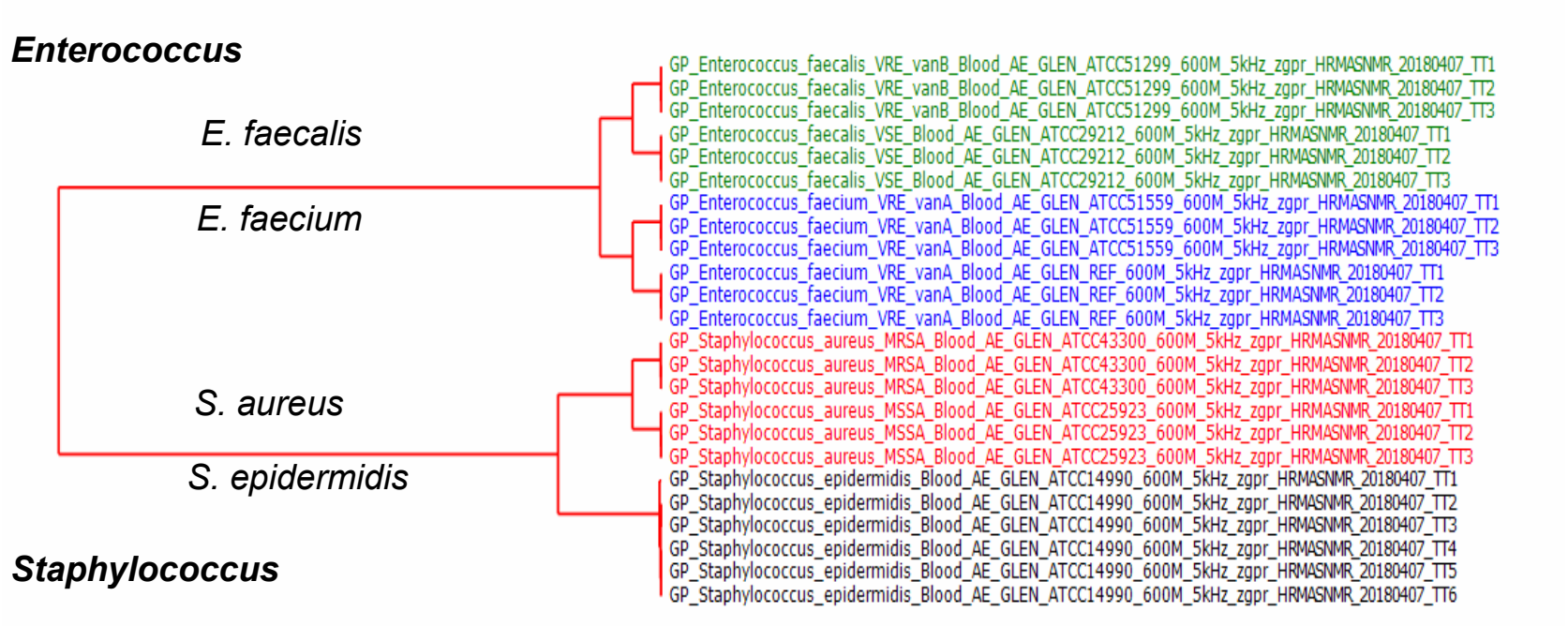
given the pulsotype label AA, B1 and CC. Feature selection algorithm was used to determine regions that contributed to successful discrimination by the PFGE pulsotypes. The ^{1}H HR-MAS NMR spectra of the three strain types, AA, B1 and CC, are shown in Figure 6.6. Spectral differences could be identified by visual inspection. A feature selection algorithm was used to determine spectral regions that contributed to discrimination among the PFGE pulsotypes. Successful discrimination by ^{1}H HR-MAS NMR spectroscopy was confirmed in the dendrogram generated by HCA using Ward linkage, resulting in 100 % concordance with PFGE pulsotypes. The variance spectrum between the ^{1}H NMR spectra of AA and CC VRE pulsotypes (Figure 6.7) observed a strong peak in the CC pulsotype isolates at δ = 3.25 ppm, which was present at much lower abundance in the AA pulsotype. Identification of the biomolecule(s) that observe ^{1}H signal at δ = 3.25 ppm may provide insight on the biochemical differences between AA and CC isolates.

Figure 6.2 Discrimination of ^{1}H HR-MAS NMR spectra acquired from yeast, and Gram-positive and Gram-negative bacteria, using spectral region δ = 0.5-4.5 ppm.



Successful discrimination between Gram-positive (GP) and Gram-negative (GN) and yeast (YT) using ^{1}H HR-MAS NMR spectra in the spectral region of δ = 0.5-4.5 ppm. Discrimination between species, *Enterococcus faecalis, E. faecium, Staphylococcus. aureus, S. epidermidis, Escherichia coli* (O157:H7 vs non-O157:H7), *Shigella flexneri* and *S. sonnei* using the same broad spectral region.

Figure 6.3 Discrimination between species, E. faecalis, E. faecium, S. aureus and S. epidermidis, demonstrated in HCA based on ^{1}H HR-MAS NMR spectra in the spectral region of δ = 0.5-4.5 ppm.



Discrimination between *Enterococcus* and *Staphylococcus*, as well as species within each genera (*E. faecium* from *E. faecalis*, and *S. aureus* from *S. epidermidis*).

Figure 6.4 HCA showing differentiation between E. coli *O157:H7, non-verotoxigenic* E. coli *and* Shigella species *grown on blood agar (A) and MacConkey agar (B) at 35 °C for 18-24 hours using region selection by* ^{1}*H HR-MAS NMR spectroscopy*

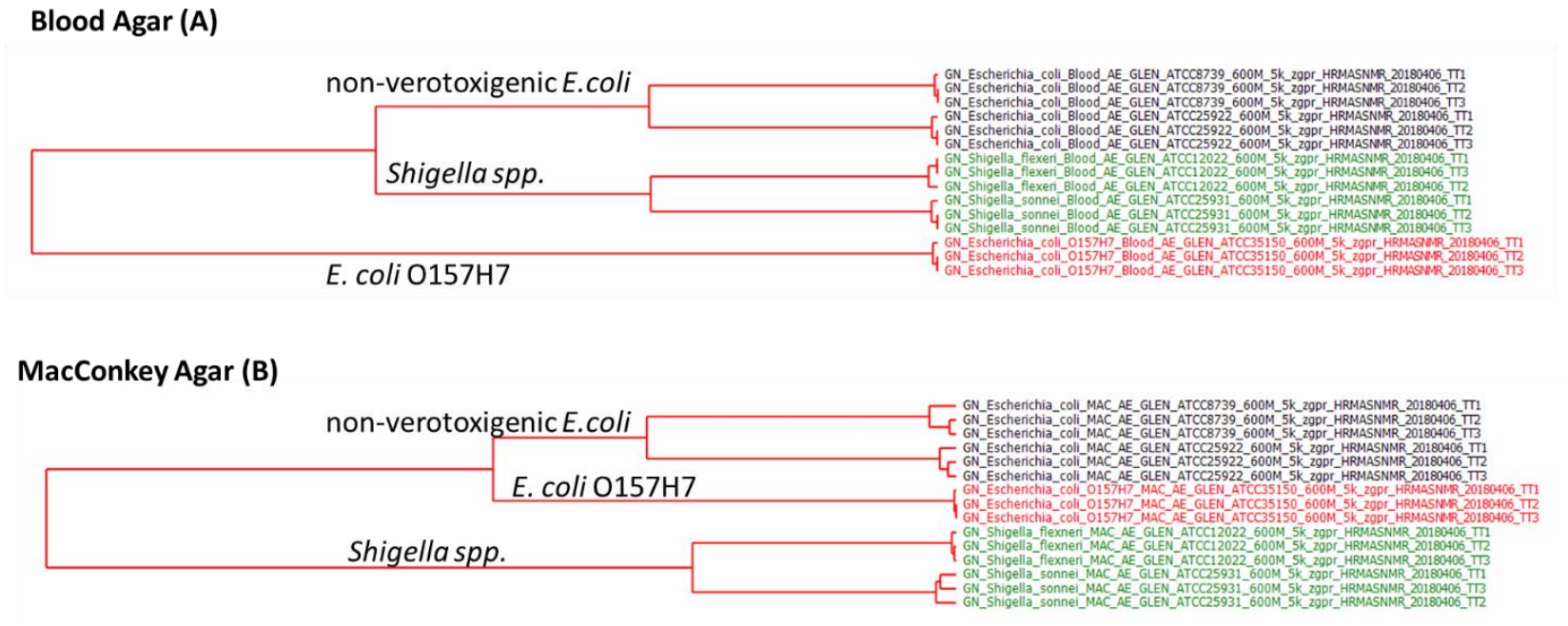


Figure 6.5 HCA showing the differentiation between MRSA (red) and MSSA (black) grown on blood agar at 35 °C for 18-24 hours using region selection by ^{1}H NMR spectroscopy

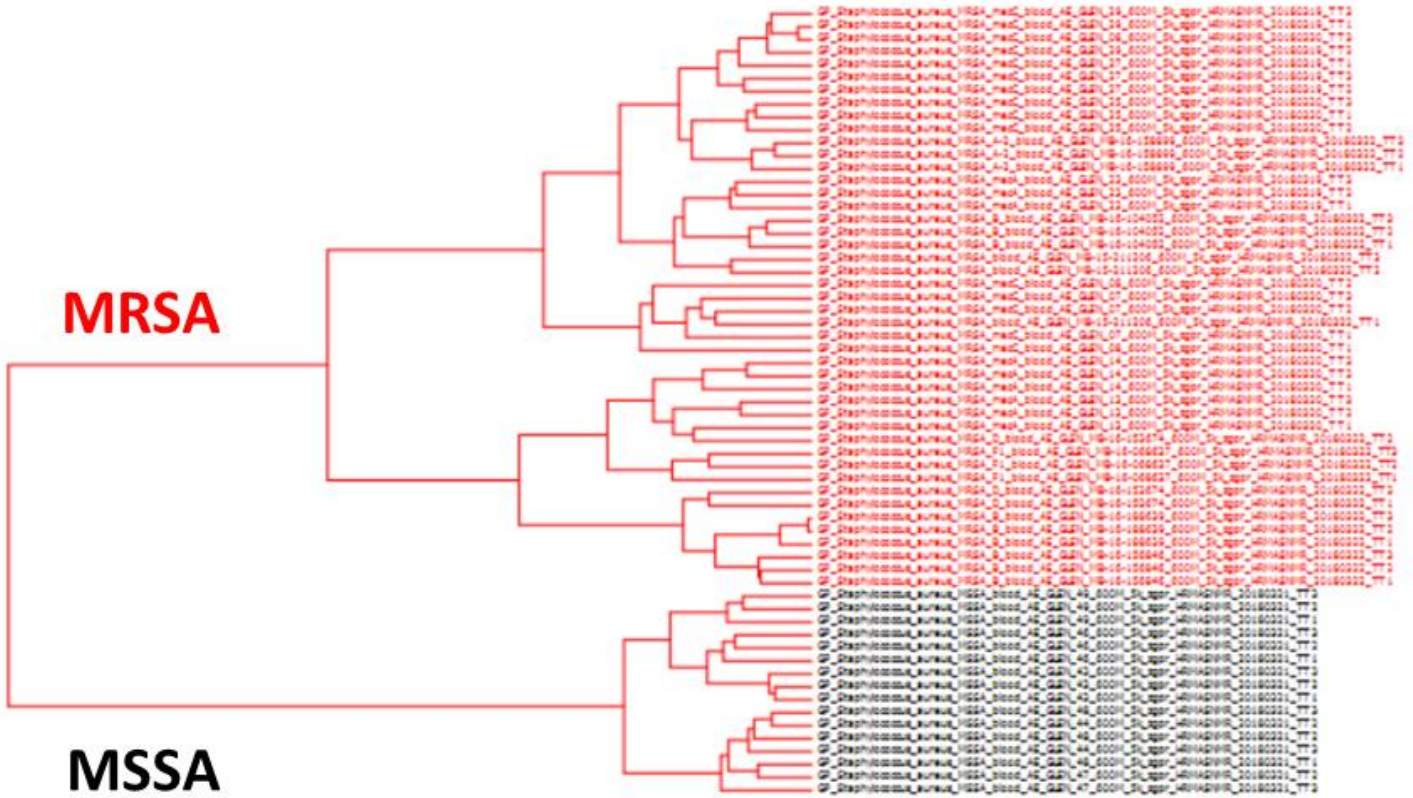
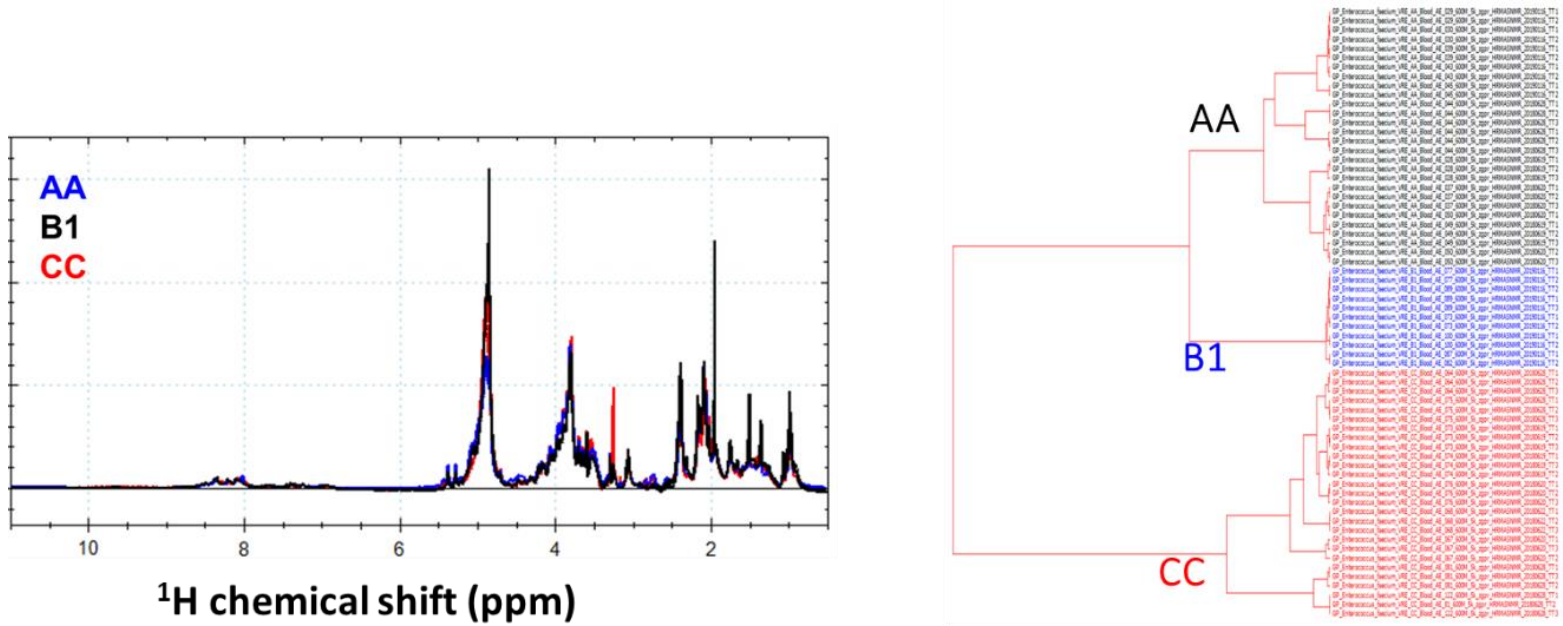
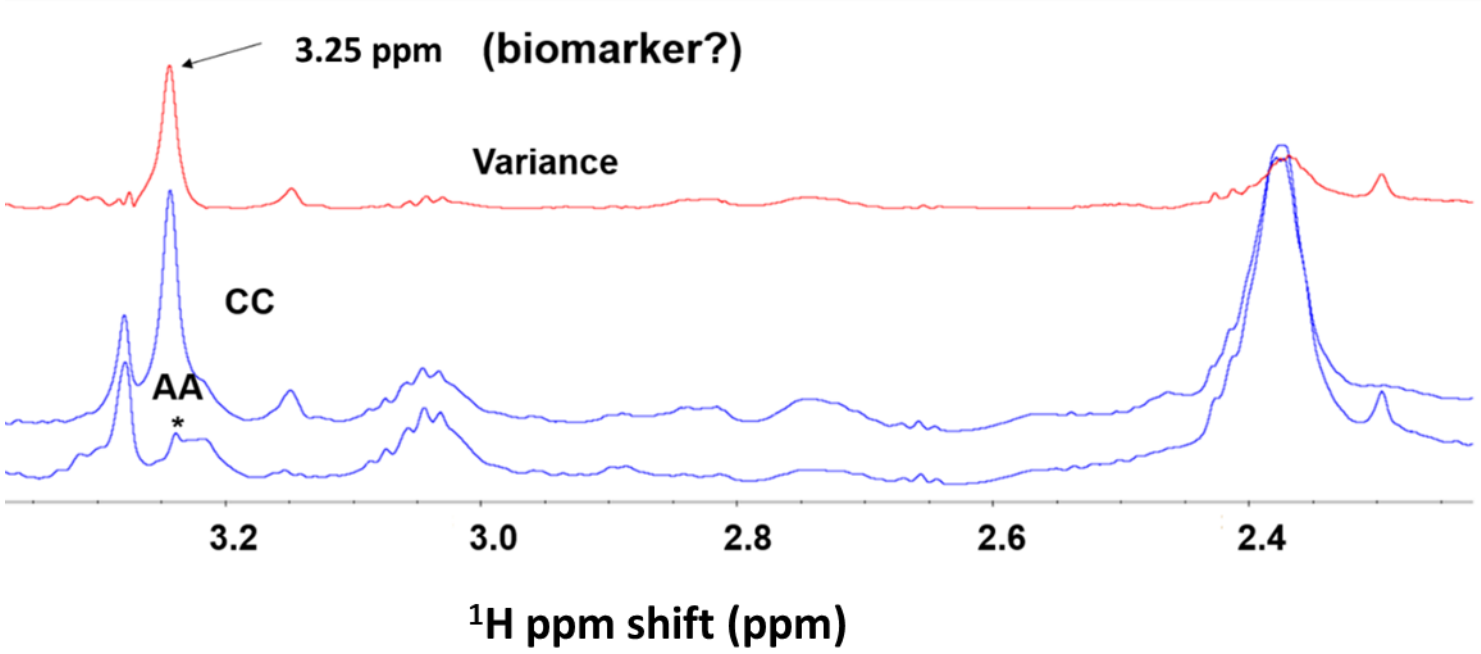


Figure 6.6 Overlaid ^{1}H HR-MAS NMR spectra of VRE isolates with pulsotypes AA, B1 and CC in the spectral range of δ = 0.5-4.5 ppm (left) and the corresponding dendrogram from HCA using Ward linkage (right)



Overlaid spectra of thre VRE pulsotypes AA (black), B1 (blue) and CC (red) in spectral region δ = 0.5-4.5 ppm. The most notable spectral difference in this region is that the CC pulsotype has an increased peak at δ = 3.25 ppm, relative to AA and B1.

Figure 6.7 Variance spectra obtained from averaged AA and CC ^{1}H chemical shift (ppm)



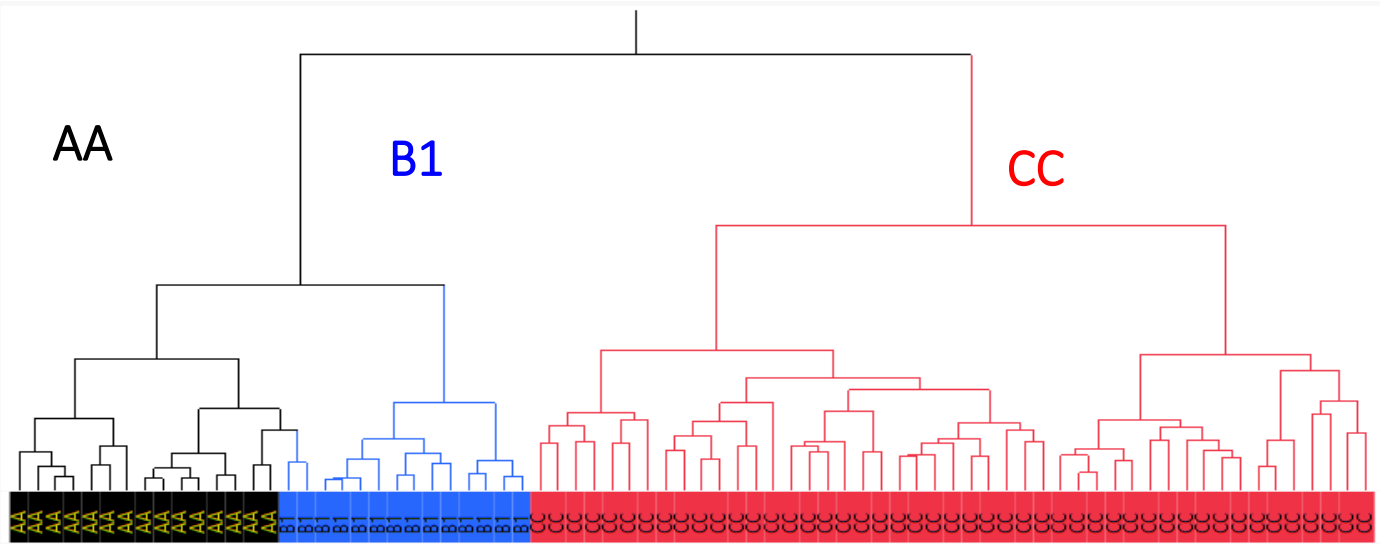
Variance spectra (red) between the AA and CC spectra (blue). Potential biomarker that differentiates the two strain types identified at $\delta = 3.25$ ppm (indicated with a star).

6.4.2 Identification of biomarker related to VRE pulsotype differences

6.4.2.1 FTIR spectral discrimination of VRE strain types

Transflection FTIR spectra were acquired from the same 28 isolates that were analyzed by ^{1}H HR-MAS NMR spectroscopy. Six replicates were acquired per isolate. Discrimination of FTIR spectra based on pulsotypes was achieved using supervised and unsupervised spectral analyses. A previously developed FTIR spectrotyping method using unsupervised PC-LDA was applied to cluster the samples based on spectral similarities. The clustering determined by the spectral similarities was then visualized in a dendrogram generated by HCA, as per Figure 6.8, which showcased concordant results with PFGE pulsotypes. With the use of a feature selection algorithm, three spectral regions that contribute to discrimination among these three PFGE pulsotypes were identified, and their associated band assignments to biomolecules are listed in Table 6.2. The selected regions were checked in the averaged second-derivative FTIR spectra (Figure 6.9), and differences were observed in the selected regions, 1054-1075, 1147-1187, and 1306-1316 cm^{-1}. Within each of the 10 to 40 cm^{-1} spectral window, peaks were identified to be associated with nucleotides, carbohydrates, and amino acids. Although there are spectral differences that correlate to PFGE pulsotypes, it was evident that biomarker identification solely based on FTIR spectroscopic data for this dataset was not possible. In the ^{1}H HR-MAS NMR spectra of VRE isolates of the AA and CC pulsotypes, the CC pulsotype exhibited a strong peak at a chemical shift correspond to the ^{1}H resonance of the methyl groups in choline (that may be part of phosphatyidylcholine) [46-49]. Consistent with the latter finding, the second derivative FTIR spectra show an increased absorbance at 970 cm^{-1}, corresponding to a band in the FTIR spectrum of choline portion of phosphatidylcholine in the CC spectrum (Figure 6.10) [50]. In order to determine if biomarkers related to identified spectral regions in transflection FTIR spectra and ^{1}H HR-MAS NMR spectra, additional spectral analyses were conducted by acquiring ^{31}P HR-MAS NMR spectra from whole cell , as well as analyses on fractionated lipophilic and hydrophilic components of one isolate each of AA and CC pulsotype VRE samples by MALDI MS, and solution state NMR spectroscopy. Whole genome sequencing data were acquired from a subset of isolates to determine whether the spectroscopic observations could be correlated to genotypic data.

Figure 6.8 Dendrogram showcasing the spectral similarities between AA, B1 and CC VRE pulsotype isolates, which were sources of three independent hospital-associated outbreaks.



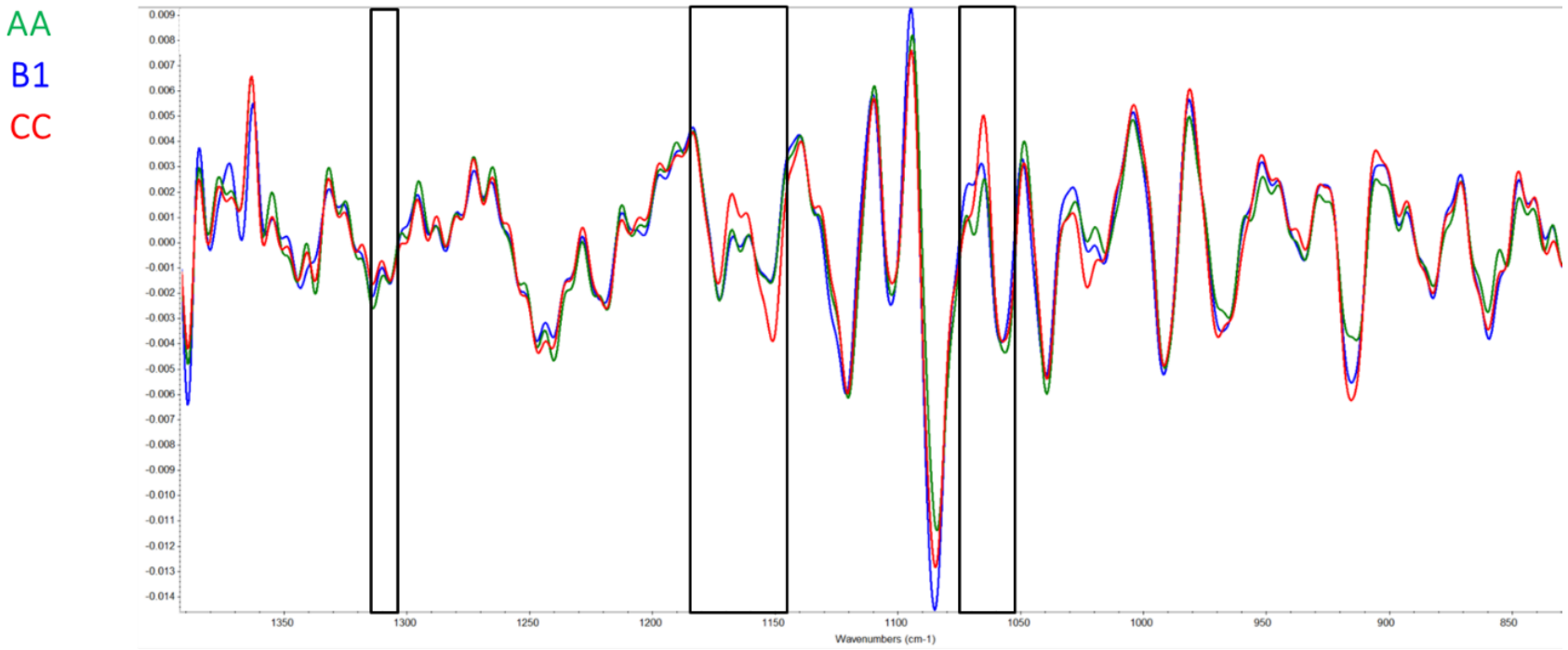
Using non-supervised spectral analyses, FTIR spectra acquired from VRE isolates with known pulsotypes AA (black), B1 (blue) and CC (red) were discriminated in concordance to the PFGE pulsotypes. The HCA used ward linkage and Euclidean distance to generate the dendrogram.

Table 6.2 Tentative band assignments associated with FTIR spectral regions that were identified by feature selection algorithm for discrimination amongst AA, B1 and CC VRE pulsotypes

Pairwise Discrimination	**Spectral Region (cm^{-1})**	**Tentative Band Assignment**	**References**
AA vs B1	1054-1075	C-OH str. membrane bound oligosaccharide PO_2- str. Phosphate residues	[51-53]
AA vs CC		C-O str. deoxyribose, ribose, phosphodiester	
AA vs CC	1147-1187	C-O str. carbohydrates, phosphodiester C-C carotenoid structure	
B1 vs CC		C-O str. from C-OH groups in serine, threonine, tyrosine C-O-C ring	
AA vs B1	1306-1316	Amide III and amino acid side chain	

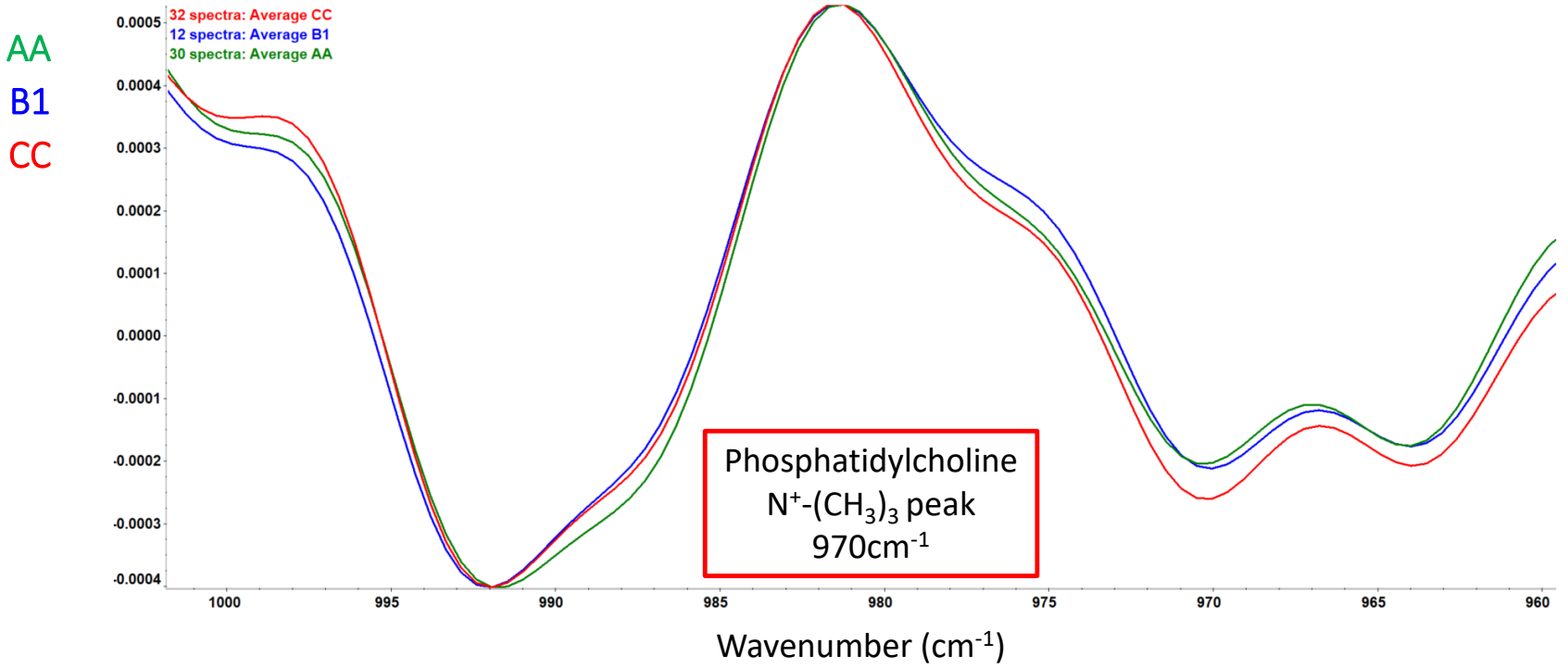
Table of spectral regions identified to discriminate between the 3 VRE pulsotypes, and their band assignments based on literature. Based on the tentative band assignments, it is clear that it is very difficult to identify specific biomarkers based on FTIR spectral regions that are different between classes of bacteria, due to overlapping absorbance signals from various biochemical components.

Figure 6.9 Second derivative FTIR spectra of VRE pulsotypes AA, B1 and CC and the spectral regions identified from feature selection algorithm display key differences that are attributed to successful spectral discrimination between the three pulsotypes.



Overlay of second derivative FTIR spectra acquired from VRE pulsotypes AA (green), B1 (blue) and CC (red). Boxed regions in the figure indicate regions that were selected using the feature selection algorithm, which identifies key regions that enhance the discrimination between labeled groups (in this case the 3 pulsotypes). Regions identified were 1054-1075 cm^{-1}, 1147-1187 cm^{-1}, and 1306-1316 cm^{-1}.

Figure 6.10 Averaged second derivative FTIR spectra of VRE pulsotypes AA, B1 and CC

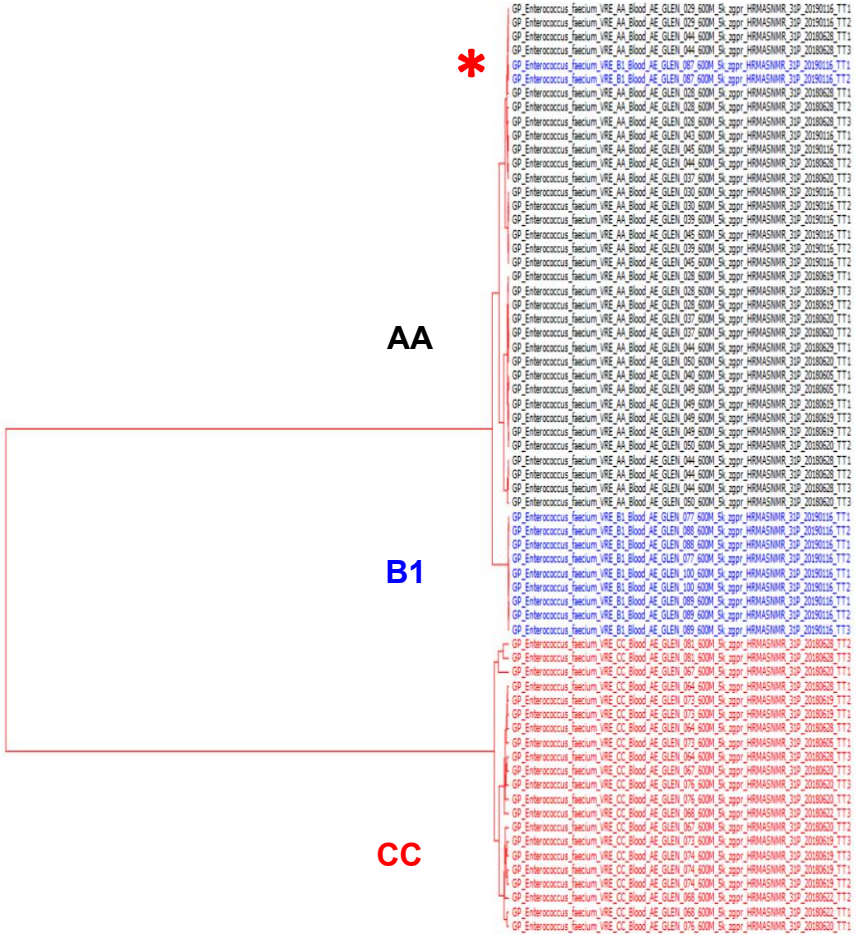


Second derivative FTIR spectra of VRE pulsotypes AA (green), B1 (blue) and CC (red) in spectral region 1000-960 cm^{-1}. CC had increased absorbance at 970 cm^{-1}, which is associated with the $N^{+}(CH_3)_3$ stretching of phosphatidylcholine, relative to other pulsotypes. This is in concordance with the increase peak observed in the 1H HR-MAS NMR spectra for CC, relative to AA and B1 pulsotype.

6.4.2.2 VRE strain typing by ^{31}P HR-MAS NMR spectroscopy

Phosphorus HR-MAS NMR spectra were acquired from VRE isolates. Due to the lower amount of ^{31}P nuclei naturally present in the samples compared to ^{1}H, the spectral acquisition required double the number of transient scans. In order to prevent interaction between the bacteria sample and the reference, an external reference (TPP dissolved in acetone) was used. TPP spectra were acquired once a day and used during spectral pre-processing to ensure that the NMR spectra were properly aligned. Spectral regions were identified using a feature selection algorithm, and clustering based on these regions was visualized in a dendrogram generated by HCA with Ward linkage as seen in Figure 6.11. Compared to the complete concordance that was achieved between ^{1}H HR-MAS NMR spectra and PFGE pulsotypes, one isolate was an outlier in the clustering based on ^{31}P HR-MAS NMR spectra, where a B1 isolate clustered with AA isolates. As in the case of the ^{1}H NMR spectra, CC pulsotype isolates had a unique ^{31}P HR-MAS NMR spectral profile that allowed them to be distinguished from AA and B1 pulsotype isolates by visual inspection of the ^{31}P NMR spectra (Figure 6.12). The differences between B1 and AA in the ^{31}P NMR spectra were subtle in comparison. The major distinguishing feature of the spectra of CC pulsotype isolates was the absence of a peak at δ = -0.82 ppm, which may be associated with phosphatidylcholine (PC), or phosphoenolypyruvic acid (PEP) (Table 6.3). In addition, the ^{31}P NMR spectra for the CC pulsotype were distinguished by the higher intensity of the peaks at δ = +0.57 ppm and δ = +0.92 ppm, which are reported to be associated with phosphodiester bridges in teichoic acid, and intracellular organic phosphate respectively [54-56].

Figure 6.11 Dendrogram generated from HCA demonstrating spectral discrimination between three VR E. faecium pulsotypes AA (Black), B1 (Blue) and CC (Red) using ^{31}P HR-MAS NMR spectra



HCA of spectra acquired by ^{31}P HR-MAS NMR spectroscopy, using regions obtained from broad region (δ = -2 to δ = 2ppm). Successful discrimination of CC from AA and B1 was achieved. One outlier of B1 clustering with AA was observed.

Figure 6.12 Stacked ^{31}P HR-MAS NMR spectra of VRE pulsotypes AA, B1 and CC in spectral range δ = -0.2 to δ = 2.5 ppm showcasing spectral differences between the three pulsotypes

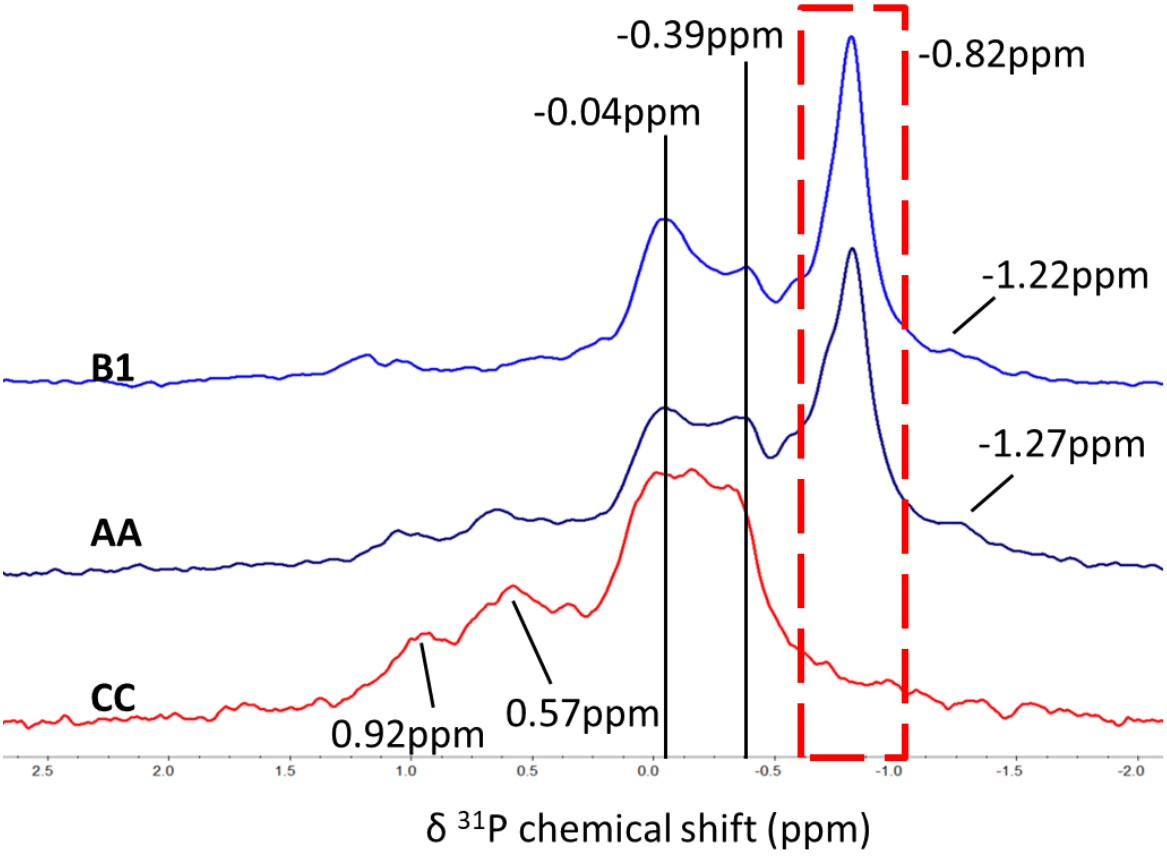


Table 6.3 Table of tentative band assignment from literature for ^{31}P NMR spectroscopy

Chemical Shift (ppm)	Tentative Band Assignment	Reference
-1.27	Nucleic acids	[53]
-0.82	Phosphoenolpyruvic acid (PEP), Phosphatidylcholine	[54]
0.00	Orthophosphoric acid	[55]
0.40-0.58	Phosphodiester bridges (in teichoic acid)	[50, 51]
0.65	Monobasic form of inorganic phosphate	[56]
0.98-0.82	Intracellular inorganic phosphate	[52]

6.4.2.3 Phenotypic antimicrobial susceptibility of VRE isolates

Isolates were tested for their antimicrobial susceptibility profile, using VITEK 2 AST, with the setting for *Enterococcus faecium*. All isolates were resistant to vancomycin (MIC ≥ 16 μg/ml), benzylpenicillin (MIC = 32 μg/ml or ≥ 64 μg/ml), ciprofloxacin (MIC ≥ 8 μg/ml), levofloxacin (MIC ≥ 8 μg/ml), and tetracycline (MIC ≥ 16 μg/ml) and had intermediate resistance to nitrofurantoin (MIC = 64 μg/ml). Table 6.4 shows results for resistance level, and MIC (μg/ml) for selected antibiotics to which the VRE isolates had varying levels of resistance. The main difference was that B1 pulsotype isolates were sensitive to the gentamicin-streptomycin combination, whereas AA and CC pulsotypes were resistant. No clear difference relating to the antimicrobial resistance profiles could be observed between AA and CC pulsotypes. For samples AA-1 and AA1 (indicated as AA_101 and AA_102 respectively in (Table 6.4), MIC for tigecycline could not be determined, even after 24hrs of run-time in the VITEK 2 AST system, and therefore was reported as terminated (TERM).

Table 6.4 Antimicrobial susceptibility test results on select antibiotics obtained from VITEK 2 AST

Code	Isolate	WGS_ID	Benzylpenicillin		Gentamicin High Level (synergy)		Streptomycin High Level (synergy)		Quinupristin/ Dalfopristin		Linezolid		Tigecycline	
AA_01	XVI360	SEQ-161	R	≥ 64	R	SYN-R	R	SYN-R	S	0.5	S	2	S	0.25
AA_02	XVI514	SEQ-162	R	≥ 64	R	SYN-R	R	SYN-R	S	0.5	S	2	S	0.25
AA_03	XVI822	SEQ-163	R	≥ 64	R	SYN-R	R	SYN-R	S	0.5	S	2	S	0.25
AA_04	XVI621	SEQ-164	R	32	R	SYN-R	R	SYN-R	S	0.5	S	2	S	≤ 0.12
AA_05	XVI633	SEQ-165	R	≥ 64	R	SYN-R	R	SYN-R	S	0.5	S	2	S	≤ 0.12
AA_06	XVI171	SEQ-166	R	≥ 64	R	SYN-R	R	SYN-R	S	0.5	S	2	S	≤ 0.12
AA_07	XVI325		R	≥ 64	R	SYN-R	R	SYN-R	S	0.5	S	2	S	≤ 0.12
AA_08	XVI264		R	32	R	SYN-R	R	SYN-R	S	0.5	S	2	S	≤ 0.12
AA_09	XVI027	SEQ-167	R	≥ 64	R	SYN-R	R	SYN-R	S	0.5	S	2	S	≤ 0.12
AA_10	XVI097		R	≥ 64	R	SYN-R	R	SYN-R	S	0.5	S	2	S	≤ 0.12
B1_01	XVII329		R	≥ 64	S	SYN-S	S	SYN-S	S	1	S	2	S	≤ 0.12
B1_02	XVII809		R	≥ 64	S	SYN-S	S	SYN-S	S	1	I	4	S	≤ 0.12
B1_03	XVII056		R	≥ 64	S	SYN-S	S	SYN-S	S	0.5	S	2	S	≤ 0.12
B1_04	XVII062	SEQ-171	R	≥ 64	S	SYN-S	S	SYN-S	S	1	S	2	S	≤ 0.12
B1_05	XVII288	SEQ-172	R	≥ 64	S	SYN-S	S	SYN-S	S	1	S	2	S	≤ 0.12
CC_01	XVII796	SEQ-173	R	≥ 64	R	SYN-R	R	SYN-R	S	0.5	S	2	S	≤ 0.12
CC_02	XVII798	SEQ-174	R	≥ 64	R	SYN-R	R	SYN-R	S	0.5	S	2	S	≤ 0.12
CC_03	XVII177	SEQ-175	R	≥ 64	R	SYN-R	R	SYN-R	S	0.5	S	2	S	≤ 0.12
CC_04	XVII293	SEQ-176	R	≥ 64	R	SYN-R	R	SYN-R	S	0.5	S	2	S	≤ 0.12
CC_05	XVII297	SEQ-177	R	≥ 64	R	SYN-R	R	SYN-R	S	0.5	S	2	S	≤ 0.12
CC_06	XVII303	SEQ-178	R	≥ 64	R	SYN-R	R	SYN-R	S	0.5	S	2	S	≤ 0.12
CC_07	XVII223	SEQ-179	R	≥ 64	R	SYN-R	R	SYN-R	S	0.5	S	2	S	≤ 0.12
CC_08	XVII599	SEQ-180	R	≥ 64	R	SYN-R	R	SYN-R	S	0.5	S	2	S	≤ 0.12
CC_09	XVII624	SEQ-181	R	≥ 64	R	SYN-R	R	SYN-R	S	0.5	S	2	S	≤ 0.12
CC_10	XVII692	SEQ-182	R	≥ 64	R	SYN-R	R	SYN-R	S	0.5	S	2	S	≤ 0.12
CC_11	XVII699	SEQ-183	R	≥ 64	R	SYN-R	R	SYN-R	S	0.5	S	2	S	≤ 0.12
CC_12	XVII747	SEQ-184	R	≥ 64	R	SYN-R	R	SYN-R	S	1	S	2	S	≤ 0.12
AA_101	XVII129	SEQ-170	R	≥ 64	R	SYN-R	R	SYN-R	S	0.5	S	2		TERM
AA_102	XVII388	SEQ-169	R	≥ 64	R	SYN-R	R	SYN-R	S	0.5	S	2		TERM

6.4.2.4 Genotypic characterization of the outbreak related VRE isolates

Whole genome sequencing data was acquired from 23 isolates using Illumina MiSeq. The quality of the data was evaluated using total base pair length, N50 and number of contigs. N_{50} is the value that describes the base pair length of the shortest contig in an assembly, where a minimum of half the total base pair length of all contigs are equal or longer in length [61]. Three isolates failed the quality check, due to low total length (1.4-1.7Mbp). The total length ranged between 2.3-2.8Mbp per isolate. The average base pairs observed by PFGE pulsotypes AA, B1 and CC were 2.53 Mbp, 2.70 Mbp, and 2.76 Mbp respectively. Total genomic length is known to vary among *E. faecium* isolates, as a result of its high genomic plasticity, through the acquisition and loss of plasmids and virulence genes [62].

Using single nucleotide polymorphism (SNP) analysis on the genome, WGS data confirmed concordance with PFGE pulsotype clustering and spectrotyping by FTIR and HR-MAS NMR spectroscopy. The cluster tree based on whole genome sequence of VRE isolates in Figure 6.13 visualized the SNP differences observed between the VRE isolates. SEQ-164 was used as the reference strain to develop the tree. Each cluster (boxed in blue, black and red) was concordant to one of the three PFGE pulsotypes identified. Numbers along the arms of the dendrogram indicate the number of SNPs between each cluster or isolate. Within each cluster, there were 0-1 SNP, showcasing high genotypic similarity among isolates in each cluster. The SNPs between clusters ranged from 32 to 58 SNP differences, showcasing clear differences between clusters. SEQ-190 and SEQ-191were *E. faecium* isolates unrelated to any of the three outbreak strain types. The SNP differences were analyzed in genes that were common across all analyzed isolates, which encompassed 23.44 % of all positions that are valid, included, and part of the core genome, indicating low similarity between the analyzed isolates. SNP analysis within each clusters was conducted using 72.26 %, 94.91 % and 85 % of all valid parts of the core genome for AA, B1 and CC pulsotype isolates respectively. The SNP differences observed (0-5 SNP) between isolates of the same pulsotype are presented in Table 6.5. The SNP analysis for all 20 isolates together only observed 27.17 % of the core genome to be common, indicating that there are large genotypic differences among the three pulsotypes. Furthermore, in the SNP analysis, 60 SNP differences were observed between AA and CC pulsotype isolates, 42-44 SNP differences between AA and B1 isolates, and 74-76 SNP differences between B1 and CC pulsotype isolates.

Whole genome sequencing data of the 20 VRE isolates were also analysed using an *in silico* MLST sequence type database from the Center for Genomic Epidemiology [42]. Based on the alleles of 7 loci related to housekeeping genes in the *Enterococcus* genome (*adk, atpA, ddl, gdh, gyd, pstS,* and *purK*), the MLST sequence types were determined. AA isolates were all identified to be ST203, whereas B1 was ST17, and CC was determined as non-typable due to the lack of the *pstS* locus. However, based on the alleles of the remaining 6 loci, the sequence type for CC pulsotype isolates was determined to be ST1478. *pstS*-null VRE *E. faecium* isolates have been observed in a number of countries like Australia, England, Scotland, South Korea and Canada [63-68]. In Canada, between 2013 and 2018 there was an increase in *pstS*-null ST1478 VRE isolates (the same ST as the CC pulsotype in this study), which coincided with increased rates of VRE bloodstream infections [64]. *pstS*-null isolates identified in other countries were of different sequence types, indicating that a unique VRE sequence type has disseminated across Canada. PstS is an inorganic phosphate binding lipoprotein that is part of the high-affinity phosphate (P_i) transport system. The lipoprotein is located in the periplasmic space, where it binds inorganic phosphate for cellular uptake [69]. High expression of the PstS protein occurs under stress, such as under alkali-acid conditions, or in the presence of sub-inhibitory concentrations of penicillin. It acts like a multi-emergency protein that help cells adapt to different habitats. The PstS protein does not exhibit high affinity for P_i when the cells have sufficient P_i [69-71].

Genes relating to antimicrobial resistance were identified for each isolate. A summary of the prevalence of the genes and their locations is presented in Table 6.7. All 20 isolates were confirmed to possess VanHAX, for vancomycin resistance. All VRE isolates' resistance genes for VanHAX were located on transposons (*Tn1656* for all AA and B1 isolates, and *Tn1656* and *Tn1649* for CC isolates). *aph(3')-IIIa* was not identified in any AA puslotype isolates. This gene encodes for APH (3')-III, an aminoglycoside phosphotransferase, and is associated with kanamycin resistance [72]. *cat* (pC221), which was only identified in CC isolates (8 of 12 isolates), is related to chloramphenicol resistance [73]. Despite tetracycline resistance observed by VITEK 2 AST in all samples, some isolates did not possess *tet(L)* nor *tet(M)* gene. While 8 out of 12 CC isolates possessed both *tet(L)* and *tet(M)* genes, 3 isolates had one of the two genes, and 1 isolate had neither. *tet(M)* was identified in 1 of 9 AA isolates, while 5 AA isolates had *tet(L)*. Sample 2020- 161 (an AA pulsotype isolate) had both genes, and 4 isolates had neither of the two genes. B1 isolates only possessed *tet(L)* genes. *msr(C),* which encodes for macrolide resistance, was

identified in 11 out of 12 CC pulsotype isolates, while 7 AA and 2 B1 isolates were found to also possess the gene. *erm(B)* was only identified on 3 out of 12 CC isolates, while all AA and B1 isolates possessed the resistance gene. AA isolates possessed the gene on *Tn1642* (*n* = 7) and on the chromosome (*n* = 2), while the gene was located on *Tn1649* in B1 isolates, and on *Tn3050* (*n* = 2) and chromosome (*n* = 1) in CC isolates. *aac(6')* was identified in all VRE isolates, all on the chromosome. This gene is intrinsic to *E. faecium,* and confers low level resistance to kanamycin and tobramycin [74]. A three-year study in Australia on *E. faecium* collected from sepsis reported 70 % of ST17 isolates (the same ST as B1) to have *aph(3')-III,* and none of the isolates possessed *cat (pC221),* while ST203 (same ST as AA) had 78 % of their isolates possessing *aph(3')-III,* and 11.1 % possessing the *cat(pC221)* gene [73]. The antimicrobial resistance patterns and genes related to antimicrobial resistance present in the VRE isolates in this study showcase that isolates belonging to the same PFGE pulsotype do not all possess the same antimicrobial resistance-related genes, nor do they necessarily possess a given resistance-related gene in the same location within their genome (i.e. chromosome vs plasmids).

VirulenceFinder was used to identify virulence factors present in all isolates [75, 78]. Sample SEQ-166 had no virulence factor identified from the assembled genome, and samples SEQ-162, SEQ-164, SEQ-180, and SEQ-181 could not have their virulence factors determined. Based on the results from VirulenceFinder, it was determined that all AA isolates had 2 virulence factors, *acm*, and efa_{Afm}, while B1 and CC isolates had 3 virulence factors, *acm*, efa_{Afm}, and hyl_{Efm} (Supplementary Table 10). Although all isolates had 100 % identity for *acm*, the genomes had varying length of the gene sequence missing (845-848 base pairs missing), with the exception of samples SEQ-171. Acm encourages collagen adhesion (cell wall-anchored collagen adhesin), and has characteristics of MSCRAMM (microbial surface components recognizing adhesive matrix molecules), and plays a role in endocarditis [79, 80]. Efa_{Afm} is an adhesion associated protein for *E. faecium*, commonly observed in clinical enterococcal isolates [81]. Hyl_{Efm} is an enzyme, glycoside hydrolase, that has β-N-acetylglucosaminidase activity, which is an important factor involved in colonization and adhesion [76, 79, 82]. Although the presence of Hyl_{Efm} does not directly contribute to *E. faecium* colonization, it has been observed that the presence of the gene hyl_{efm} resulted in higher bacteria load during colonization [82]. CC pulsotype isolates were found to have more plasmids and virulence genes compared to AA and B1 pulsotype isolates, which correlates with the larger genome size observed for CC isolates.

PlasmidFinder was used to determine the presence of plasmids in the VRE isolates (Supplementary Table 11) [41, 75]. Plasmids make up a large part of the accessory genes in *Enterococcus* genomes and are responsible for horizontal gene transfer of antibiotic resistance and virulence characteristics of hospital adapted strains [76]. Four to nine plasmids were identified per isolate. Overall, the median number of plasmids per PFGE pulsotype were 7, 6, and 8 for AA, B1 and CC respectively. AA, B1 and CC pulsotypes had plasmids from rep_trans, rep1, rep3, and repA_N family. Only CC pulsotype isolates had plasmids from the Inc18 family. Plasmid pNB2354 is known to have association with *hly* adhesion [77]; however, not all VRE isolates investigated in the current study had *hly* genes identified as a virulence factor. Plasmid pEF418 was only identified in AA and B1 isolates, and pKQ10 was only found in AA isolates. Plasmid pRE25 is known to carry antimicrobial resistance genes such as *erm(B), cat* and *aph(3')-III* [77]. This plasmid was only identified in CC pulsotype isolates, and the associated antimicrobial resistance genes were identified.

Figure 6.13 Whole genome sequencing cluster tree of VRE isolates originating from three different outbreaks

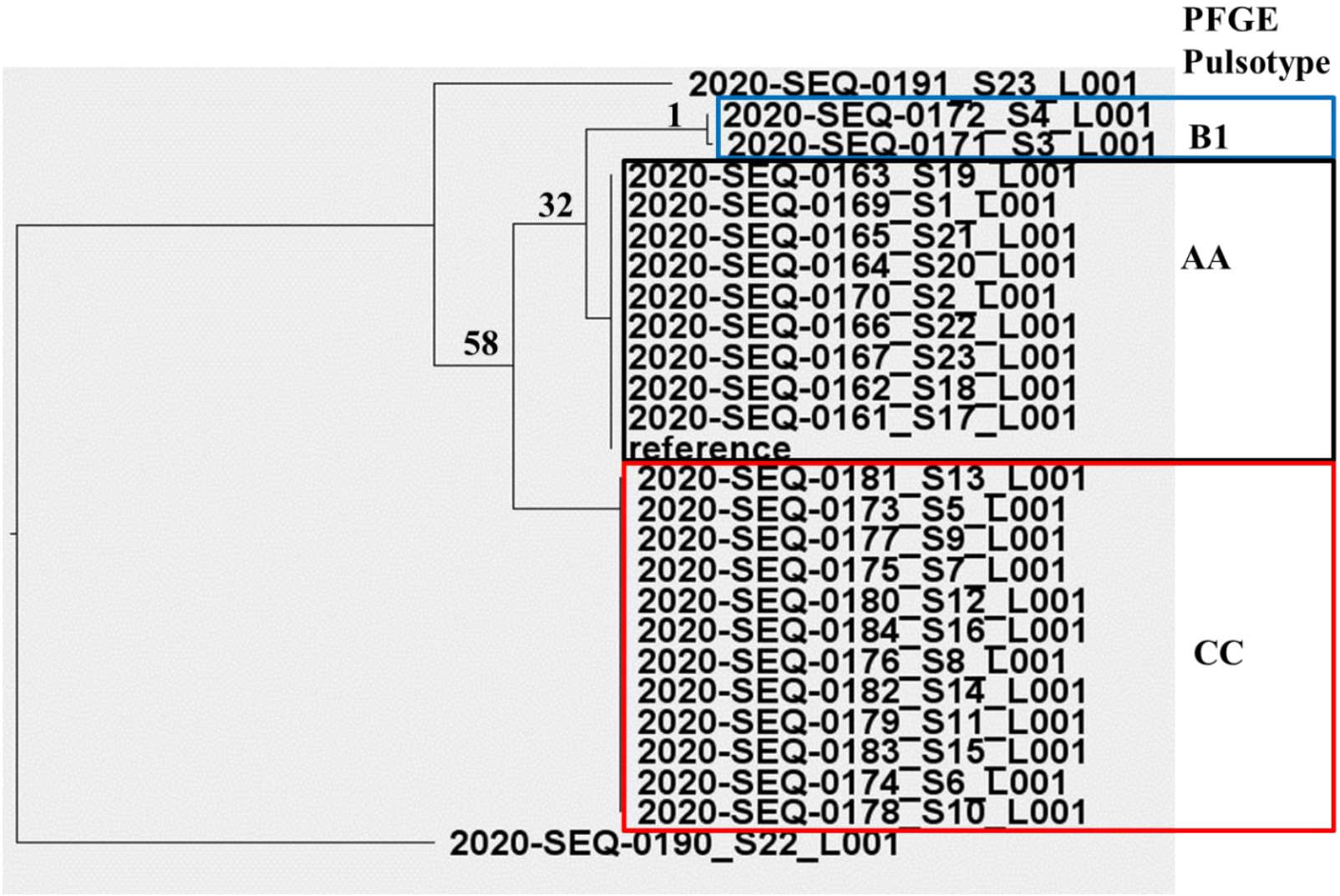


Table 6.5 The percentage of core genome utilized in SNP analysis per VRE pulsotype cluster, and the SNP difference observed between samples

Pulsotype	No. of samples analyzed	% core genome utilized in analysis	SNP difference between samples
AA	7	72.26 %	1-5
B1	2	4.91 %	2
CC	11	85 %	0-1
All	20	27.17 %	AA-B1: 42-44 AA-CC: 60 B1-CC: 74-76

Table 6.6 MLST sequence type identified from WGS data using in silico MLST sequence analyses

PFGE	Sequence Type
AA	203
B1	17
CC	Non-typable → 1478

Figure 6.14 Example of MLST result for a CC pulsotye VRE isolate

MLST-2.0 Server - Results

mlst Profile: *efaecium*

Organism: *Enterococcus faecium*

Sequence Type: *Unknown*

Nearest STs: *117,889,1651,1478,1465,677,721,780,1587,1518*

Locus	Identity	Coverage	Alignment Length	Allele Length	Gaps	Allele
adk	100.0	100.0	437	437	0	adk_1
atpA	100.0	100.0	556	556	0	atpA_9
ddl	100.0	100.0	465	465	0	ddl_1
gdh	100.0	100.0	530	530	0	gdh_1
gyd	100.0	100.0	395	395	0	gyd_1
pstS						No hit found
purK	100.0	100.0	492	492	0	purK_1

extended output

Input Files: *2020-SEQ-0175.fasta*

Table 6.7 Resistance genes and their locations identified in the VRE isolates

Gene	Allele	Resistance	Location	AA		B1		CC	
aac(6')	1	aminoglycoside	Chromosome	7	(77.8 %)	2	(100 %)	11	(91.7 %)
ant(6)-la	1	aminoglycoside	Plasmid 1642	7	(77.8 %)				
ant(6)-la	1	aminoglycoside	Plasmid 1656	1	(11.1 %)				
ant(6)-la	1	aminoglycoside	Plasmid 1649			2	(100 %)		
ant(6)-la	1	aminoglycoside	Chromosome					1	(8.3 %)
ant(6)-la	1	aminoglycoside	Plasmid 3050					9	(75 %)
ant(6)-la	1	aminoglycoside	Plasmid 1866					1	(8.3 %)
ant(6)-la	1	aminoglycoside	Plasmid 1656					1	(8.3 %)
aph(3')-III	1	aminoglycoside	Chromosome					1	(8.3 %)
aph(3')-III	1	aminoglycoside	Plasmid 1649			2	(100 %)		
aph(3')-III	1	aminoglycoside	Plasmid 1656					9	(75 %)
aph(3')-III	1	aminoglycoside	Plasmid 1866					1	(8.3 %)
aph(3')-III	1	aminoglycoside	Plasmid 3050					1	(8.3 %)
cat(pC221)	1	phenicol	Chromosome					3	(25 %)
cat(pC221)	1	phenicol	Novel_0					3	(25 %)
cat(pC221)	1	phenicol	Plasmid 1866					1	(8.3 %)
cat(pC221)	1	phenicol	Plasmid 2351					1	(8.3 %)
dfrG	1	trimethoprim	Chromosome	7	(77.8 %)	2	(100 %)	8	(66.7 %)
erm(B)	7	macrolide	Chromosome	1	(11.1 %)				
erm(B)	9	macrolide	Chromosome	1	(11.1 %)				
erm(B)	9	macrolide	Plasmid 1642	7	(77.8 %)				
erm(B)	9	macrolide	Plasmid 1649			2	(100 %)		
erm(B)	9	macrolide	Plasmid 3050					2	(16.7 %)
msr(C)	1	macrolide	Chromosome	7	(77.8 %)	2	(100 %)	11	(91.7 %)
tet(L)	2	tetracycline	Chromosome					1	(8.3 %)
tet(L)	2	tetracycline	Plasmid 2351	6	(66.7 %)	2	(100 %)	6	(50 %)
tet(L)	2	tetracycline	Plasmid 1866	1	(11.1 %)			2	(16.7 %)
tet(M)	10	tetracycline	Chromosome					1	(8.3 %)
tet(M)	10	tetracycline	Plasmid 1866	1	(11.1 %)			9	(75 %)
VanHAX	2	glycopeptide	Novel_1					1	(8.3 %)
VanHAX	2	glycopeptide	Chromosome	1	(11.1 %)				
VanHAX	2	glycopeptide	Plasmid 1649					6	(50 %)
VanHAX	2	glycopeptide	Plasmid 1656	8	(88.9 %)	2	(100 %)	5	(41.7 %)

6.4.2.5 Preliminary results from NMR spectroscopy and MALDI MS on fractionated components

One isolate each of AA and CC VRE pulsotype were subjected to extractions for separation and analysis of hydrophilic and hydrophobic components, using a modified Folch method. The extraction was done using deuterated solvents, in order to acquire NMR spectra in solution state directly from the hydrophilic and hydrophobic phases. The aqueous phase contained molecules that are highly polar, and miscible in D_2O and deuterated methanol. The non-aqueous phase contained molecules that were miscible in deuterated chloroform. All insoluble components were not analysed in this study. To ensure that there were extracted metabolites and biomolecules in both the hydrophilic and hydrophobic phases, 3-10 µl of each phase was deposited onto a low E-glass slide for spectral acquisition by transflection FTIR microspectroscopy, using a 250 µm aperture microscope. FTIR spectra indicated that there were more variance between AA and CC in the hydrophobic phase compared to the hydrophilic phase (data not shown).

6.4.2.5.1 Solution-state NMR spectroscopy

Proton and ^{31}P NMR spectra were acquired from both D_2O/methanol-D and chloroform-D phases, for both AA and CC VRE isolates using a 600-MHz NMR spectrometer (Bruker, GE). The ^{31}P NMR spectra were externally referenced, using TPP dissolved in acetone. Proton NMR spectra were referenced using methanol-D peaks at δ = 3.35 ppm and δ = 4.78 ppm. Proton NMR spectra acquired from CC samples were scaled based on the difference in biomass of AA and CC (0.38 g and 0.60 g, respectively) at the extraction step. The ^{1}H NMR spectra acquired from the D_2O/methanol-D phase of the AA and CC extracts are shown in Figure 6.15. The CC spectrum had higher relative intensity at δ = 1.90 ppm, δ = 2.00 ppm, and δ = 2.47-2.67 ppm; and exhibited a multiplet at δ = 4.40-4.44 ppm whereas the AA spectrum showed higher relative peaks between δ = 2.01-2.06 ppm, as well as peaks at δ = 2.1-2.18 ppm and a triplet at δ = 2.36-2.40 ppm. Thus, these spectra are indicative of differences in the metabolite profiles of the AA and CC pulsotype isolates. In addition, the intense peak at δ = 2.00 ppm in the CC spectrum may be associated with unsaturated lipid, as the extraction used methanol, and it is possible that the aqueous phase may contain highly polar lipids. In previous studies employing the Folch extraction method, phospholipids such as PC, sphingomyelin, and cardiolipin have been observed in both the chloroform and the methanol-aqueous phase, and some phospholipids such as lysophospholipids are more polar and have lower solubility in chloroform than others and tend to partition into the aqueous phase [83]. However, the possibility that differences in phospholipid composition

between AA and CC isolates may have contributed to differences in the ^{1}H NMR spectra of the aqueous phase obtained by extraction of these isolates is not supported by the ^{31}P NMR spectra acquired from the aqueous phase, which show little difference between AA and CC samples (Figure 6.16). Additional ^{31}P NMR experiments should be conducted in the future, with different relaxation times to ensure that signals from molecules of various sizes and mobilities are observed.

For the chloroform phase of the extraction, ^{31}P NMR spectra were processed with line broadening (5 Hz) as an additional spectral pre-processing step (Figure 6.17). Peaks were identified in the CC spectrum at δ = -1.11 ppm and δ = +0.45 ppm. There were no discernible peaks in the AA spectrum. Phosphatidylglycerol (PG) was reported to be observed at δ = 0.43-0.47 ppm in the ^{31}P NMR spectra [84] and thus the peak observed at δ = 0.45 ppm in the ^{31}P NMR spectrum from the CC extracted sample, may tentatively be assigned to PG. Proton NMR spectra showed little difference between the AA and CC samples (Figure 6.18, only showing regions with visible differences in spectra). The CC spectrum had a doublet at δ = 4.75 and δ = 4.76 ppm that was not present in the spectrum of the AA sample.

6.4.2.5.2 MALDI mass spectrometry

MALDI MS data were acquired from the D_2O/ deuterated methanol and deuterated chloroform extracted components using an MS imaging system. There were no peaks observed in positive and negative modes using 1,5-diaminonaphthalene (DAN) matrix (data not shown), while peaks were observed in the D_2O/ deuterated methanol phase of both AA (black) and CC (blue) samples using 1, 5-dihydroxybenzoic acid (DHB) matrix (Figure 6.19). These peaks did not match the profiles of lipids, and therefore are likely derived from small molecules, peptides or sugars that are part of the metabolic or structural components of AA and CC VRE. These observed peaks could not be identified at the current moment, however, in the future, further analyses using tandem MS may allow for identification.

From the chloroform extracts of AA and CC samples, MALDI MS spectra were also acquired in triplicate deposition, using DAN matrix in negative and positive modes. In the positive mode, peaks related to phosphatidylcholine and fatty acyl-CoA were identified in both AA and CC samples. CC observed higher intensity peaks compared to AA, however, after taking into consideration the initial biomass used for extraction, the difference was not statistically significant for both PC (Table 6.8) and fatty acyl-CoA (Table 6.9). In the negative mode, MS peaks observed

for CC (red) showed overall upregulation of phospholipids, specifically phosphatidylinositol (PI) and phosphatidic acid (PA) or phosphatidylglycerol (PG), compared to AA (blue) (Figure 6.20). Even after normalization, adjusting for the difference in biomass during extraction, CC had statistically significantly higher levels of these phospholipids compared to AA (Table 6.10).

Figure 6.15 Overlay of ^{1}H NMR spectra in region 0.6-2.8 ppm (top) and 4.2-5.4ppm (bottom) acquired from D_2O-methanol-D phase of extraction from AA (red) and CC (blue) VRE isolates

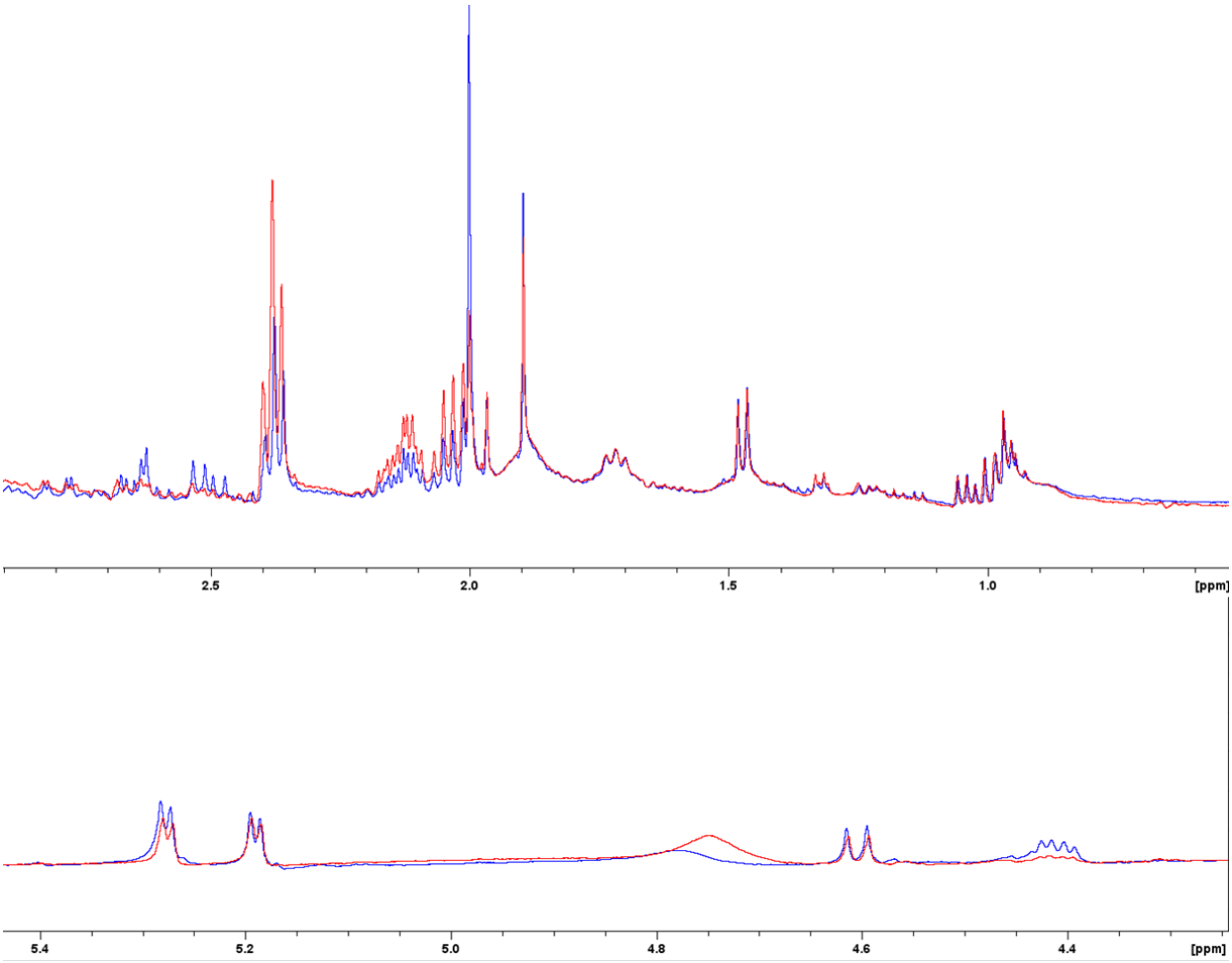


Figure 6.16 Overlay of 31P NMR spectra acquired from D2O-methanol-D extracted portions of AA (red) and CC (blue) VRE isolates

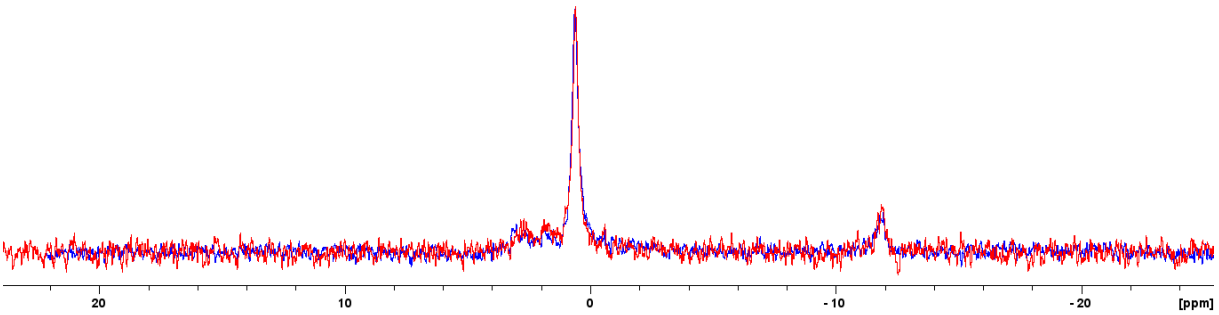


Figure 6.17 Overlay of ^{31}P NMR spectra from chloroform-phase of extraction obtained from AA (red) and CC (blue) VRE isolates

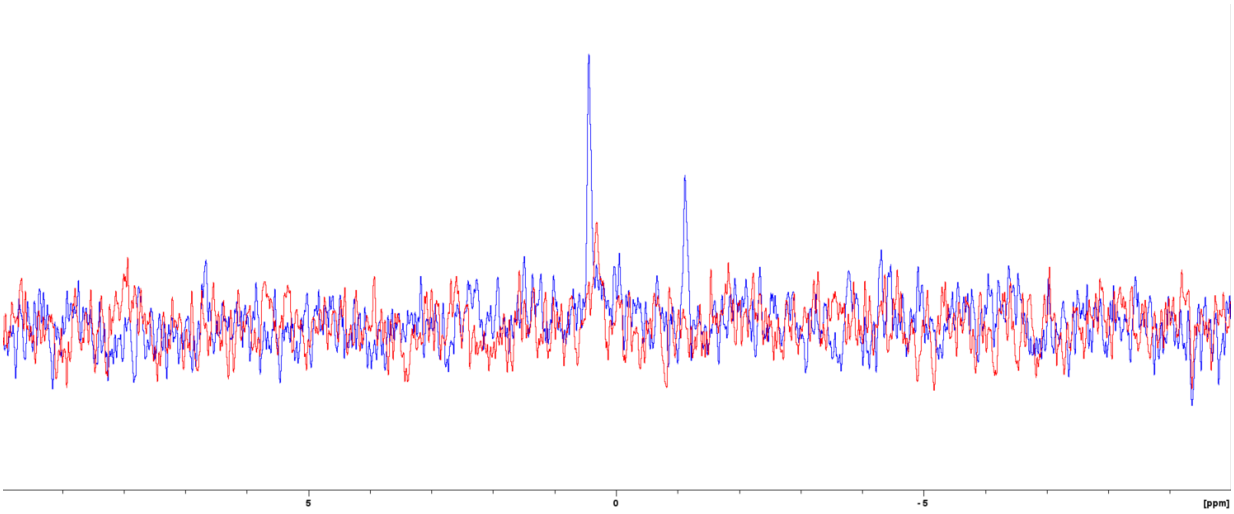


Figure 6.18 Overlay of ^{1}H NMR spectra in the region δ = 3.9-5.0ppm for chloroform phase of AA (red) and CC (blue) VRE isolates

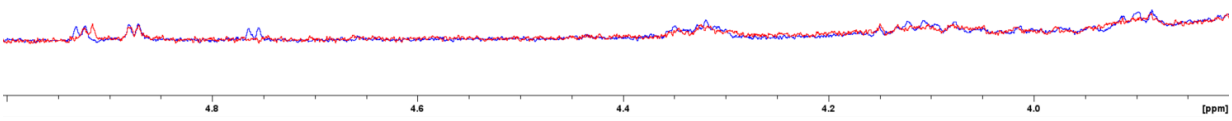


Figure 6.19 MS peaks obtained from aqueous phase extraction from AA (black) and CC (blue) VRE samples. Red spectrum represents peaks of the DAN matrix

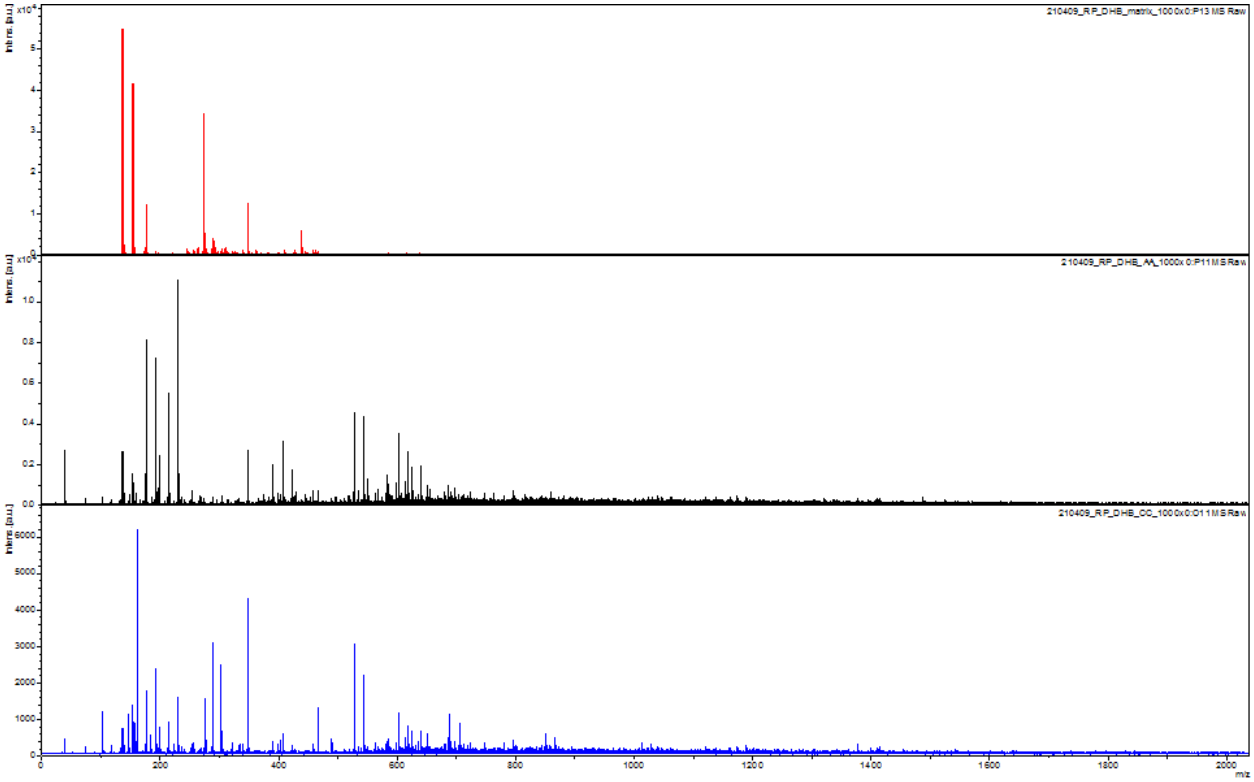


Table 6.8 Statistical Analysis on the peak heights at centroids related to phosphatidylcholine, obtained from positive mode analysis on chloroform extraction of AA and CC VRE Isolates

	AA			**CC**		
Centroid [m/z]	**Intensity**	**SD**	**Variance**	**Intensity**	**SD**	**Variance**
756.56	0.472	0.349	0.122	0.736	0.284	0.081
758.57	2.061	1.451	2.104	3.215	1.190	1.417
760.59	1.371	0.896	0.803	2.034	0.736	0.542
780.56	1.006	0.733	0.537	1.460	0.510	0.260
782.57	0.869	0.612	0.375	1.401	0.502	0.252
784.59	1.043	0.776	0.602	1.568	0.609	0.371
786.59	1.782	1.230	1.512	2.475	0.932	0.868
788.62	0.800	0.513	0.263	1.031	0.363	0.132
802.54	0.122	0.069	0.005	0.207	0.064	0.004
804.56	0.135	0.077	0.006	0.203	0.067	0.005
806.58	0.545	0.410	0.168	0.743	0.287	0.082
808.61	1.270	0.963	0.927	1.915	0.734	0.538
810.63	1.119	0.821	0.675	1.540	0.619	0.383
812.62	0.213	0.146	0.021	0.278	0.112	0.012
830.61	0.154	0.093	0.009	0.241	0.080	0.006
832.61	0.191	0.129	0.017	0.265	0.096	0.009
834.64	0.206	0.152	0.023	0.261	0.110	0.012
836.66	0.359	0.272	0.074	0.431	0.188	0.035
Sum No Norm.	13.72	2.87	8.24	20.00	2.24	5.01
mass (g)	0.38			0.6		
Sum Norm.	36.10	7.56		33.34	3.73	
p Value	0.601					
Result	**Not significant**					

Table 6.9 Statistical Analysis on peak heights at centroids related to fatty acyl-CoAs from positive mode analysis on chloroform extractions from AA and CC VRE Isolates

	AA			**CC**		
Centroid [m/z]	**Intensity**	**SD**	**Variance**	**Intensity**	**SD**	**Variance**
940.022	0.219	0.139	0.019	0.426	0.213	0.046
942.102	0.389	0.259	0.067	0.734	0.362	0.131
988.423	0.070	0.065	0.004	0.110	0.070	0.005
Sum No Norm.	0.68	0.30	0.09	1.27	0.43	0.18
mass (g)	0.38			0.6		
Sum Norm.	1.78	0.79		2.12	0.71	
p Value	0.609					
Result	**Not significant**					

Figure 6.20 Overlaid mass spectra acquired from chloroform extracts from VRE AA and CC pulsotypes in region 600-1400 m/z in negative mode

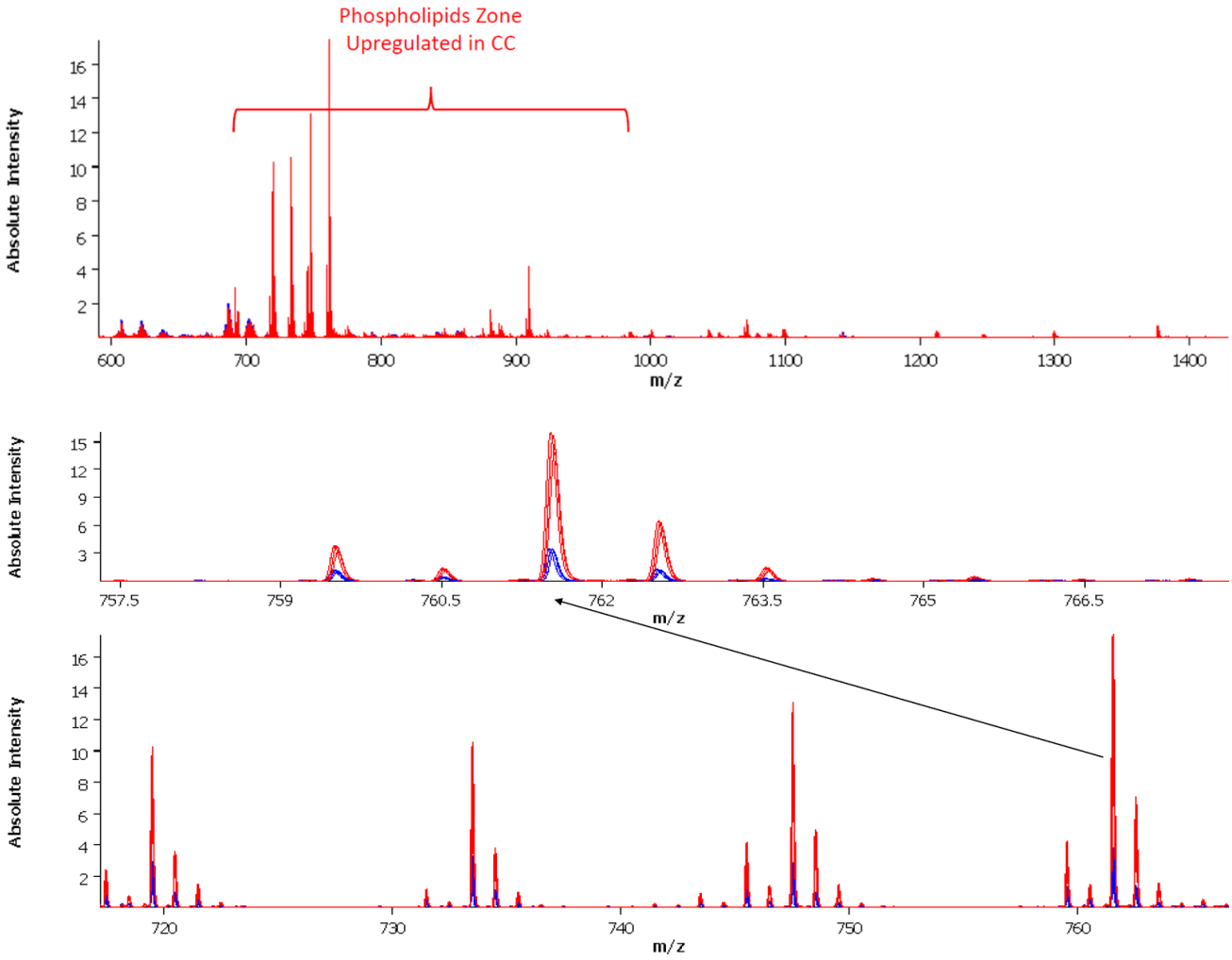


Table 6.10 Statistical analysis on centroids related to phospholipids acquired from negative mode analysis from chloroform extracts of AA and CC VRE Isolates

	AA			**CC**		
Centroid [m/z]	**Intensity**	**STD**	**Variance**	**Intensity**	**STD**	**Variance**
719.553	2.573	1.767	3.121	9.339	3.678	13.526
733.513	2.990	1.943	3.777	9.551	3.572	12.762
747.527	2.525	1.683	2.833	11.979	4.378	19.164
761.562	3.491	2.215	4.906	15.932	5.575	31.082
909.513	0.057	0.032	0.001	3.629	1.674	2.804
1069.584	0.155	0.127	0.016	0.558	0.325	0.105
1071.544	0.240	0.184	0.034	0.941	0.486	0.236
Sum No Norm.	12.03	3.83	14.69	51.93	8.93	79.68
mass (g)	0.38			0.6		
Sum Norm.	31.66	10.09		86.55	14.88	
p Value	0.0061					
Result	**Significant**					

6.5 Discussion

Whole organism analyses of VRE isolates belonging to three PFGE pulsotypes (denoted as AA, B1 and CC) by ^{1}H HR-MAS NMR and FTIR spectroscopy enabled discrimination among the pulsotypes, demonstrating the presence of spectrally measurable biochemical differences among the pulsotypes. The possibility of identifying the nature of these biochemical differences beginning with the discovery of biomarkers was investigated through an examination of the spectral differences among the pulsotypes. Choline-containing compounds emerged as candidate biomarkers of the CC pulsotype with the observation that spectral features in both the ^{1}H NMR and FTIR spectra could be consistent with higher levels of choline in the CC pulsotype isolates. However, ^{31}P HR-MAS NMR spectra of the VRE isolates revealed that a distinguishing feature of the CC pulsotype was the absence of a peak observed in the spectra of isolates belonging to the other pulsotypes at a chemical shift value corresponding to the ^{31}P NMR resonance of phosphatidylcholine. Although PC was observed in larger amounts from CC hydrophilic extract, the amount was not significant, and corresponding peaks were not observed in the ^{31}P NMR spectra. Although glycerophospholipid metabolism in *E. faecium* has been mapped in KEGG and is shown to produce PC, PC has not been reported to be present as a cell-membrane constituent in *E. faecium*, nor in most Gram-positive bacteria [85-88]. Based on the FTIR and ^{1}H HR-MAS NMR spectra, choline itself (not bound to any phosphate) remains a potential candidate as a biomarker differentiating between AA and CC isolates, but cannot be confirmed with the current results. Choline is known to be taken up by many bacteria as a precursor for betaine and glycine betaine, which are osmoprotectants, and aids in survival in presence of varying osmotic pressures [89-91].

Mass spectra obtained from the chloroform extracts of AA and CC pulsotype isolates indicated significantly increased levels of PI and PA/PG in CC isolates, which was in concordance with the peak observed in the ^{31}P NMR spectra. PA has a peak at $\delta = 0.23$ ppm, which was not observed, and therefore, the mass spectra peaks are likely of PG [84]. PG is one of the most common phospholipids found in *E. faecium* along with lysyl-PG (LPG), cardiolipin (CL) and glycreophospho-diglycodiacylglycerol [86]. LPG is a modified PG through aminoacylation, which makes the phospholipid charge cationic or zwitterionic. In addition to LPG, alanine-PG, arginine-PG, lysine-CL and alanine-CL have been found in *E. faecium* [86, 92]. Triple specific aaPGS (arginine, alanine, lysine) gene is known to allow for aminoacylation of phospholipids in *E. faecium*, and is known to be important for its adaptation to acidic pH [93]. Modified PGs are related

to cellular homeostasis and pathogenicity, through increased bacterial resistance to antibiotics, bacteriocins and host defense molecules, resulting in increased virulence [86]. Approximate phospholipid composition of *E. faecium* is 44 % CL, 23 % PG, and 32 % amino-containing phospholipid, with a large portion of the amino-containing phospholipid being lysyl-phosphatidyl-glycerol [94-96]. The peak in the ^{31}P NMR spectrum at δ = -1.1 ppm could not be assigned to a compound. PC is reported to have peaks around δ = -0.84 ppm [84], and therefore it is unlikely that the peak at δ = -1.1 ppm is from PC. Since the ^{31}P NMR peaks of polar lipids are known to shift with variables such as pH, concentration of lipids, and chain length of the fatty acids, comparison and peak assignments are generally much more difficult for ^{31}P NMR than ^{1}H NMR [97]. Alterations in phospholipids impair cellular envelope structure and function, biofilm formation, fitness, susceptibility to environmental stresses that are dependent on phospholipid [98].

The sample preparation prior to ^{31}P NMR spectra on the extracted components was a critical limitation in this study. Upon extracting the lipid soluble components, chloroform solvent was not desiccated for reconstitution at a higher concentration for increased signal. Furthermore, addition of chelating agent such as EDTA in $CDCl_3$ should be added for chelation of paramagnetic, divalent cations. EDTA reduces line broadening which are caused by the binding of these cations to negatively charged phospholipids [97]. Another alternative sample preparation for further analyses is to dissolve the phospholipids in aqueous sodium cholate at pH = 7.5 [97]. As for the ^{1}H NMR spectra from the extracts, it is possible that signals observed in regions that overlap with the methanol-D and D_2O may have been lost. Additionally, the mixture of D_2O and methanol-D solvents (likely at different ratios between AA and CC isolates) would have affected the chemical shifts of the compounds. Therefore, a more controlled sample preparation is required when obtaining NMR spectra from extracted compounds that represent many more samples of both AA and CC isolates.

Furthermore, isolated of cell wall components should be studied in whole-cell and extracted components, to determine whether capsular polysaccharide structures of *E. faecium* isolates vary by sequence types. *E. faecium* possess *epa* (enterococcal polysaccharide antigen) locus that produce capsular polysaccharides and proteins on the cell wall surface, which increase their virulence, and adhesion capabilities [99]. Capsular polysaccharide structure of *E. faecium* is not well-studied yet, however, it was reported to consist of a 6-α-D-glucose-1-2-glycerol-3-phosphate backbone with a substitution on the C-2 of glucose with an α-2-1-D-glucose residue [99].

Hendrickx et al (2013) suggested that *E. faecium* isolates are capable of producing distinct types of capsular polysaccharides on the cell surface, due to the presence of a novel capsule 'cps' like region in the genome, that is also involved in the phosphoregularotry system [99]. Multidimensional NMR experiments on whole-cell and cell wall components may provide additional information on these cell wall and capsular polysaccharide components, all of which have phosphate and polysaccharide components [99]. In the present study, analysis of the mass spectra acquired from fractionated portions was focused on the identification of phospholipids. However, additional analyses and spectral search on small metabolites observed from the hydrophilic portion of the cell fractionations may provide insight on whether differences in (phosphorylated-)polysaccharide presence exist between the AA and CC isolates. Little is known about the *pstS*-null ST1478, or the effects of the lack of the *pstS* locus. PstS binds phosphates for cellular uptake and is involved in two pathways, ABC transporters and the two-component system. According to known pathways described in KEGG, the lack of PstS can be compensated in phosphate uptake by utilizing other proteins like SenX3 and PhoA, under a phosphate limited condition [87, 88]. Additional metabolomic analyses on CC isolates with and without modification for *pstS* gene insertion could aid in understanding the role that the gene plays in *E. faecium*, as well as its effects on the structural and metabolic characteristics of the cells. Based on the large differences observed in the core and pangenome of AA, B1 and CC pulsotypes, further analysis of the genomic data would need to focus on correlating specific biomolecular differences that were observed and understanding the genes and pathways involved. Long-read whole genome sequencing is currently in progress to obtain a better understanding of the VRE genome. The data obtained from long-read WGS will enable for higher confidence in the genomic differences that can be found between the strain types, and also make it easier to identify the specific SNP locations that are related to functions and phenotypic characteristics relating to the differences observed in NMR and FTIR spectra. The spectroscopic and genotypic analyses all revealed differences between the VRE outbreak strains, but definite correlation between spectroscopic and genotypic data could not be achieved at this time.

6.6 Conclusion

In this study, ^{1}H and ^{31}P HR-MAS NMR spectra were acquired from live bacteria for a limited number of isolates belonging to different genera and species to demonstrate that there are sufficient spectral features that enable for successful spectral discrimination. This was further demonstrated with successful discrimination of VRE isolates in concordance with the PFGE pulsotypes, AA, B1 and CC. Taking advantage of the higher chemical specificity of NMR spectroscopy in comparison to FTIR spectroscopy, identification of biomarkers relating to successful spectral discrimination between the VRE strain types was attempted. Potential biomarkers were identified from the ^{1}H and ^{31}P HR-MAS NMR spectra. This was analysed in combination with WGS data, and NMR and MS data from fractionated hydrophilic and lipophilic portions of AA and CC VRE isolates. With the WGS data, differences between pulsotypes were identified in the virulence genes and plasmids possessed, as well as in the MLST sequence type. In particular, the lack of the *pstS* gene in CC pulsotype VRE isolates was an interesting discovery. Although differences in phosphorous-containing molecules within AA and CC pulsotype isolates were observed, it was not clear whether there was a direct correlation between the lack of the *pstS* gene and spectral differences observed by FTIR and NMR spectroscopy. Spectroscopic analyses suggested that phosphorous-containing molecules such as phospholipids, specifically PC, PI, and PA/PG were present in different proportions in AA and CC pulsotypes. To expand on the preliminary work conducted in this study, the identification of biomolecules contributing to successful FTIR spectral discrimination at the sub-species level for VRE isolates should be further investigated using tandem mass spectrometry and multidimensional NMR experiments on the non-soluble components, such as the cell wall.

.

6.7 References

1. Bağcıoğlu, M., et al., *Detection and identification of Bacillus cereus, Bacillus cytotoxicus, Bacillus thuringiensis, Bacillus mycoides and Bacillus weihenstephanensis via machine learning based FTIR Spectroscopy.* Frontiers in microbiology, 2019. **10**: p. 902.
2. Dieckmann, R., et al., *Rapid characterisation of Klebsiella oxytoca isolates from contaminated liquid hand soap using mass spectrometry, FTIR and Raman spectroscopy.* Faraday discussions, 2016. **187**: p. 353-375.
3. Dziuba, B., *Identification of selected Leuconostoc species with the use of FTIR spectroscopy and artificial neural networks.* Acta Scientiarum Polonorum Technologia Alimentaria, 2011. **10**(3): p. 275-285.
4. Essendoubi, M., et al., *Epidemiological investigation and typing of Candida glabrata clinical isolates by FTIR spectroscopy.* Journal of microbiological methods, 2007. **71**(3): p. 325-331.
5. Lamprell, H., et al., *Discrimination of Staphylococcus aureus strains from different species of Staphylococcus using Fourier transform infrared (FTIR) spectroscopy.* International journal of food microbiology, 2006. **108**(1): p. 125-129.
6. Lecellier, A., et al., *Implementation of an FTIR spectral library of 486 filamentous fungi strains for rapid identification of molds.* Food microbiology, 2015. **45**: p. 126-134.
7. Sousa, C., et al., *Development of a FTIR-ATR based model for typing clinically relevant Acinetobacter baumannii clones belonging to ST98, ST103, ST208 and ST218.* Journal of Photochemistry and Photobiology B: Biology, 2014. **133**: p. 108-114.
8. Tsutsumi, T., et al. *2125. Staphylococcus Species Identification by Fourier Transform Infrared (FTIR) Spectroscopic Techniques: A Cross-Lab Study*. in *Open Forum Infectious Diseases*. 2019. Oxford University Press US.
9. Vogt, S., et al., *Fourier-transform infrared (FTIR) spectroscopy for typing of clinical Enterobacter cloacae complex isolates.* Frontiers in microbiology, 2019. **10**.
10. Wenning, M., et al., *Identification and differentiation of food-related bacteria: a comparison of FTIR spectroscopy and MALDI-TOF mass spectrometry.* Journal of microbiological methods, 2014. **103**: p. 44-52.
11. Lam, L.M., et al., *Reagent-Free Identification of Clinical Yeasts Using Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy.* Journal of clinical microbiology, 2019: p. JCM. 01739-18.
12. Lam, T.M.L., *Rapid Identification and Classification of Escherichia Coli and Shigella by Attenuated Total Reflectance-Fourier Transform Infrared Spectroscopy*. 2017, McGill University Libraries.
13. Vallieres, E., et al. *Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy for Rapid Identification of Non-Fermenting Gram-Negative Bacilli Isolated from Patients with Cystic Fibrosis*. in *Open forum infectious diseases*. 2017. Oxford University Press.
14. Markley, J.L., et al., *The future of NMR-based metabolomics.* Current opinion in biotechnology, 2017. **43**: p. 34-40.
15. Wong, A., C. Boutin, and P.M. Aguiar, *1H high resolution magic-angle coil spinning (HR-MACS) μNMR metabolic profiling of whole Saccharomyces cervisiae cells: a demonstrative study.* Frontiers in chemistry, 2014. **2**: p. 38.

16. Chauton, M.S., et al., *HR MAS 1H NMR spectroscopy analysis of marine microalgal whole cells.* Marine Ecology Progress Series, 2003. **256**: p. 57-62.
17. Valentini, M., et al., *The HRMAS–NMR tool in foodstuff characterisation.* Magnetic Resonance in Chemistry, 2011. **49**: p. S121-S125.
18. Wang, Y., et al., *Magic Angle Spinning NMR and 1H− 31P Heteronuclear Statistical Total Correlation Spectroscopy of Intact Human Gut Biopsies.* Analytical chemistry, 2008. **80**(4): p. 1058-1066.
19. Li, W., et al., *Methods for acquisition and assignment of multidimensional high-resolution magic angle spinning NMR of whole cell bacteria.* Analytical chemistry, 2005. **77**(18): p. 5785-5792.
20. Li, W., *Multidimensional HRMAS NMR: a platform for in vivo studies using intact bacterial cells.* Analyst, 2006. **131**(7): p. 777-781.
21. Klinowski, J., *Magic-angle-spinning NMR.* Solid State Ionics, 1985. **16**: p. 3-14.
22. Bourne, R., et al., *Identification of Enterococcus, Streptococcus, andStaphylococcus by Multivariate Analysis of Proton Magnetic Resonance Spectroscopic Data from Plate Cultures.* Journal of clinical microbiology, 2001. **39**(8): p. 2916-2923.
23. Himmelreich, U., et al., *A rapid screening test to distinguish between Candida albicans and Candida dubliniensis using NMR spectroscopy.* FEMS microbiology letters, 2005. **251**(2): p. 327-332.
24. Himmelreich, U., et al., *Rapid identification of Candida species by using nuclear magnetic resonance spectroscopy and a statistical classification strategy.* Appl. Environ. Microbiol., 2003. **69**(8): p. 4566-4574.
25. Sorrell, T.C., et al., *Application of proton nuclear magnetic resonance spectroscopy to the study of Cryptococcus and cryptococcosis.* FEMS yeast research, 2006. **6**(4): p. 558-566.
26. Gudlavalleti, S.K., et al., *In vivo determination of Neisseria meningitidis serogroup A capsular polysaccharide by whole cell high-resolution magic angle spinning NMR spectroscopy.* Carbohydrate research, 2006. **341**(4): p. 557-562.
27. Jachymek, W., et al., *Structures of the O-specific polysaccharides from Yokenella regensburgei (Koserella trabulsii) strains PCM 2476, 2477, 2478, and 2494: high-resolution magic-angle spinning NMR investigation of the O-specific polysaccharides in native lipopolysaccharides and directly on the surface of living bacteria.* Biochemistry, 1999. **38**(36): p. 11788-11795.
28. Lee, R.E., et al., *Rapid structural characterization of the arabinogalactan and lipoarabinomannan in live mycobacterial cells using 2D and 3D HR-MAS NMR: structural changes in the arabinan due to ethambutol treatment and gene mutation are observed.* Glycobiology, 2004. **15**(2): p. 139-151.
29. Maes, E., et al., *Molecular phenotyping of mannosyltransferases-deficient Candida albicans cells by high-resolution magic angle spinning NMR.* Journal of biochemistry, 2009. **145**(4): p. 413-419.
30. Zhou, X. and L. Cegelski, *Nutrient-dependent structural changes in S. aureus peptidoglycan revealed by solid-state NMR spectroscopy.* Biochemistry, 2012. **51**(41): p. 8143-8153.
31. Romaniuk, J.A. and L. Cegelski, *Peptidoglycan and Teichoic Acid Levels and Alterations in S. aureus by Cell-Wall and Whole-Cell NMR.* Biochemistry, 2018.

32. Romaniuk, J.A. and L. Cegelski, *Bacterial cell wall composition and the influence of antibiotics by cell-wall and whole-cell NMR.* Phil. Trans. R. Soc. B, 2015. **370**(1679): p. 20150024.
33. Szymanski, C.M., et al., *Detection of conserved N-linked glycans and phase variable lipo-oligosaccharides and capsules from Campylobacter cells by mass spectrometry and high resolution magic angle spinning NMR spectroscopy.* Journal of Biological Chemistry, 2003.
34. Bingol, K. and R. Brüschweiler, *Two elephants in the room: new hybrid nuclear magnetic resonance and mass spectrometry approaches for metabolomics.* Current opinion in clinical nutrition and metabolic care, 2015. **18**(5): p. 471.
35. Zhou, X., et al., *Enterococcus faecium: from microbiological insights to practical recommendations for infection control and diagnostics.* Antimicrobial Resistance & Infection Control, 2020. **9**(1): p. 1-13.
36. AlMasoud, N., et al., *Rapid discrimination of Enterococcus faecium strains using phenotypic analytical techniques.* Analytical Methods, 2016. **8**(42): p. 7603-7613.
37. Bortolaia, V., et al., *ResFinder 4.0 for predictions of phenotypes from genotypes.* Journal of Antimicrobial Chemotherapy, 2020. **75**(12): p. 3491-3500.
38. Joensen, K.G., et al., *Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli.* Journal of clinical microbiology, 2014. **52**(5): p. 1501-1510.
39. Tetzschner, A.M.M., et al., *In silico genotyping of Escherichia coli isolates for extraintestinal virulence genes by use of whole-genome sequencing data.* Journal of clinical microbiology, 2020. **58**(10).
40. Carattoli, A., et al., *In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing.* Antimicrobial agents and chemotherapy, 2014. **58**(7): p. 3895-3903.
41. Carattoli, A., et al., *PlasmidFinder and pMLST: in silico detection and typing of plasmids.* Antimicrobial agents and chemotherapy, 2014.
42. Larsen, M.V., et al., *Multilocus sequence typing of total-genome-sequenced bacteria.* Journal of clinical microbiology, 2012. **50**(4): p. 1355-1361.
43. Lévesque, S., et al., *A side by side comparison of Bruker Biotyper and VITEK MS: utility of MALDI-TOF MS technology for microorganism identification in a public health reference laboratory.* PloS one, 2015. **10**(12): p. e0144878.
44. Himmelreich, U., et al., *Rapid etiological classification of meningitis by NMR spectroscopy based on metabolite profiles and host response.* PloS one, 2009. **4**(4): p. e5328.
45. Himmelreich, U., T.C. Sorrell, and H.-M. Daniel, *Nuclear magnetic resonance spectroscopy-based identification of yeast*, in *Human Fungal Pathogen Identification*. 2017, Springer. p. 289-304.
46. García-García, A.B., et al., *1H HR-MAS NMR-based metabolomics analysis for dry-fermented sausage characterization.* Food chemistry, 2018. **240**: p. 514-523.
47. Pagter, M., C.C. Yde, and K.H. Kjær, *Metabolic fingerprinting of dormant and active flower primordia of ribes nigrum using high-resolution magic angle spinning NMR.* Journal of agricultural and food chemistry, 2017. **65**(46): p. 10123-10130.

48. Bundy, J.G., et al., *Discrimination of pathogenic clinical isolates and laboratory strains of Bacillus cereus by NMR-based metabolomic profiling.* FEMS microbiology letters, 2005. **242**(1): p. 127-136.
49. Duarte, I.F., et al., *Analytical approaches toward successful human cell metabolome studies by NMR spectroscopy.* Analytical chemistry, 2009. **81**(12): p. 5023-5032.
50. Derenne, A., et al., *Infrared Spectroscopy of Membrane Lipids*, in *Encyclopedia of Biophysics*, G.C.K. Roberts, Editor. 2013, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 1074-1081.
51. da Cunha, B.R., et al., *Fourier-Transform Mid-Infrared Spectroscopy in Biomedicine.* Essential Techniques for Medical and Life Scientists: A guide to contemporary methods and current applications with the protocols: Part 2, 2020: p. 1.
52. Naumann, D., *FT-infrared and FT-Raman spectroscopy in biomedical research.* Applied spectroscopy reviews, 2001. **36**(2-3): p. 239-298.
53. Lu, X., et al., *Application of mid-infrared and Raman spectroscopy to the study of bacteria.* Food and Bioprocess Technology, 2011. **4**(6): p. 919-935.
54. Ezra, F.S., et al., *Phosphorus-31 and carbon-13 nuclear magnetic resonance studies of anaerobic glucose metabolism and lactate transport in Staphylococcus aureus cells.* Biochemistry, 1983. **22**(16): p. 3841-3849.
55. Bychowska, A., et al., *Chemical structure of wall teichoic acid isolated from Enterococcus faecium strain U0317.* Carbohydrate research, 2011. **346**(17): p. 2816-2819.
56. Carvalho, J., et al., *Development of a bioreactor system for cytotoxic evaluation of pharmacological compounds in living cells using NMR spectroscopy.* Journal of pharmacological and toxicological methods, 2019. **95**: p. 70-78.
57. Mandala, V.S., et al., *Bacterial Phosphate Granules Contain Cyclic Polyphosphates: Evidence from 31P Solid-State NMR.* Journal of the American Chemical Society, 2020.
58. Burt, C.T., S.M. Cohen, and M. Bárány, *Analysis of intact tissue with 31P NMR.* Annual review of biophysics and bioengineering, 1979. **8**(1): p. 1-25.
59. Lipok, J., et al., *Prospects of in vivo 31P NMR method in glyphosate degradation studies in whole cell system.* Enzyme and Microbial Technology, 2009. **44**(1): p. 11-16.
60. Slonczewski, J.L., et al., *pH homeostasis in Escherichia coli: measurement by 31P nuclear magnetic resonance of methylphosphonate and phosphate.* Proceedings of the National Academy of Sciences, 1981. **78**(10): p. 6271-6275.
61. Quainoo, S., et al., *Whole-genome sequencing of bacterial pathogens: the future of nosocomial outbreak analysis.* Clinical microbiology reviews, 2017. **30**(4): p. 1015-1063.
62. Arredondo-Alonso, S., et al., *Genomes of a major nosocomial pathogen Enterococcus faecium are shaped by adaptive evolution of the chromosome and plasmidome.* BioRxiv, 2019: p. 530725.
63. Lemonidis, K., et al., *Missing pstS locus associated with an insertion in tetM in vancomycin resistant Enterococcus faecium.* bioRxiv, 2017: p. 236786.
64. McCracken, M., et al., *Emergence of pstS-Null Vancomycin-Resistant Enterococcus faecium Clone ST1478, Canada, 2013–2018.* Emerging Infectious Diseases, 2020. **26**(9): p. 2247.
65. Lemonidis, K., et al., *Emergence of an Australian-like pstS-null vancomycin resistant Enterococcus faecium clone in Scotland.* Plos one, 2019. **14**(6): p. e0218185.

66. Kim, H.M., et al., *Emergence of vancomycin-resistant Enterococcus faecium ST1421 lacking the pstS gene in Korea.* European Journal of Clinical Microbiology & Infectious Diseases, 2020. **39**(7): p. 1349-1356.
67. Raven, K.E., et al., *A decade of genomic history for healthcare-associated Enterococcus faecium in the United Kingdom and Ireland.* Genome research, 2016. **26**(10): p. 1388-1396.
68. Carter, G.P., et al., *Emergence of endemic MLST non-typeable vancomycin-resistant Enterococcus faecium.* Journal of Antimicrobial Chemotherapy, 2016. **71**(12): p. 3367-3371.
69. O'May, G., et al., *The high-affinity phosphate transporter Pst in Proteus mirabilis HI4320 and its importance in biofilm formation.* Microbiology, 2009. **155**(Pt 5): p. 1523.
70. Esteban, A., et al., *Expression of the pstS gene of Streptomyces lividans is regulated by the carbon source and is partially independent of the PhoP regulator.* BMC microbiology, 2008. **8**(1): p. 1-12.
71. Allenby, N.E., et al., *Post-transcriptional regulation of the Bacillus subtilis pst operon encoding a phosphate-specific ABC transporter.* Microbiology, 2004. **150**(8): p. 2619-2628.
72. Chow, J.W., *Aminoglycoside resistance in enterococci.* Clinical Infectious Diseases, 2000. **31**(2): p. 586-589.
73. Lee, T., et al., *A three-year whole genome sequencing perspective of Enterococcus faecium sepsis in Australia.* PloS one, 2020. **15**(2): p. e0228781.
74. Hollenbeck, B.L. and L.B. Rice, *Intrinsic and acquired resistance mechanisms in enterococcus.* Virulence, 2012. **3**(5): p. 421-569.
75. Camacho, C., et al., *BLAST+: architecture and applications.* BMC bioinformatics, 2009. **10**(1): p. 1-9.
76. Santagati, M., F. Campanile, and S. Stefani, *Genomic diversification of enterococci in hosts: the role of the mobilome.* Frontiers in microbiology, 2012. **3**: p. 95.
77. Gilmore, M.S., et al., *Enterococci: From Commensals to Leading Causes of Drug Resistant Infection [Internet].* 2014.
78. Clausen, P.T., F.M. Aarestrup, and O. Lund, *Rapid and precise alignment of raw reads against redundant databases with KMA.* BMC bioinformatics, 2018. **19**(1): p. 1-8.
79. Weber, A., et al., *Increase of vancomycin-resistant Enterococcus faecium strain type ST117 CT71 at Charité-Universitätsmedizin Berlin, 2008 to 2018.* Antimicrobial Resistance & Infection Control, 2020. **9**(1): p. 1-9.
80. Freitas, A.R., et al., *Multidrug-resistant high-risk Enterococcus faecium clones: can we really define them?* International Journal of Antimicrobial Agents, 2020: p. 106227.
81. Soheili, S., et al., *Wide distribution of virulence genes among Enterococcus faecium and Enterococcus faecalis clinical isolates.* The Scientific World Journal, 2014. **2014**.
82. Gao, W., B.P. Howden, and T.P. Stinear, *Evolution of virulence in Enterococcus faecium, a hospital-adapted opportunistic pathogen.* Current opinion in microbiology, 2018. **41**: p. 76-82.
83. Ahmmed, M.K., et al., *Simple and Efficient One-Pot Extraction Method for Phospholipidomic Profiling of Total Oil and Lecithin by Phosphorus-31 Nuclear Magnetic Resonance Measurements.* Journal of Agricultural and Food Chemistry, 2020. **68**(48): p. 14286-14296.

84. Kaffarnik, S., et al., *Two-dimensional 31P, 1H NMR spectroscopic profiling of phospholipids in cheese and fish.* Journal of agricultural and food chemistry, 2013. **61**(29): p. 7061-7069.
85. Sohlenkamp, C., I.M. López-Lara, and O. Geiger, *Biosynthesis of phosphatidylcholine in bacteria.* Progress in lipid research, 2003. **42**(2): p. 115-162.
86. Mishra, N.N., et al., *Daptomycin resistance in enterococci is associated with distinct alterations of cell membrane phospholipid content.* PLoS One, 2012. **7**(8): p. e43958.
87. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes.* Nucleic acids research, 2000. **28**(1): p. 27-30.
88. Kanehisa, M., *Toward understanding the origin and evolution of cellular organisms.* Protein Science, 2019. **28**(11): p. 1947-1951.
89. Heo, S., et al., *Genomic Insight into the Salt Tolerance of Enterococcus faecium, Enterococcus faecalis and Tetragenococcus halophilus.* Journal of microbiology and biotechnology, 2019. **29**(10): p. 1591-1602.
90. Boch, J., B. Kempf, and E. Bremer, *Osmoregulation in Bacillus subtilis: synthesis of the osmoprotectant glycine betaine from exogenously provided choline.* Journal of Bacteriology, 1994. **176**(17): p. 5364-5371.
91. Boncompagni, E., et al., *Occurrence of choline and glycine betaine uptake and metabolism in the family Rhizobiaceae and their roles in osmoprotection.* Applied and environmental microbiology, 1999. **65**(5): p. 2072-2077.
92. Slavetinsky, C., S. Kuhn, and A. Peschel, *Bacterial aminoacyl phospholipids–Biosynthesis and role in basic cellular processes and pathogenicity.* Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids, 2017. **1862**(11): p. 1310-1318.
93. Smith, A.M., et al., *A conserved hydrolase responsible for the cleavage of aminoacylphosphatidylglycerol in the membrane of Enterococcus faecium.* Journal of Biological Chemistry, 2013. **288**(31): p. 22768-22776.
94. Epand, R.M. and R.F. Epand, *Lipid domains in bacterial membranes and the action of antimicrobial agents.* Biochimica et Biophysica Acta (BBA)-Biomembranes, 2009. **1788**(1): p. 289-294.
95. Sakayori, Y., et al., *Characterization of Enterococcus faecium mutants resistant to mundticin KS, a class IIa bacteriocin.* Microbiology, 2003. **149**(10): p. 2901-2908.
96. Duitschaever, C. and D. Jordan, *Development of resistance to heat and sodium chloride in Streptococcus faecium recovering from thermal injury.* Journal of Milk and Food Technology, 1974. **37**(7): p. 382-386.
97. Schiller, J., et al., *31P NMR spectroscopy of phospholipids: From micelles to membranes.* Current Analytical Chemistry, 2007. **3**(4): p. 283-301.
98. Rowlett, V.W., et al., *Impact of membrane phospholipid alterations in Escherichia coli on cellular function and bacterial stress adaptation.* Journal of bacteriology, 2017. **199**(13).
99. Hendrickx, A.P., W. Van Schaik, and R.J. Willems, *The cell wall architecture of Enterococcus faecium: from resistance to pathogenesis.* Future Microbiology, 2013. **8**(8): p. 993-1010.

Chapter 7. General Discussion

Current strain type characterization used to supplement epidemiological information for outbreak detection relies on the use of PFGE. This long and laborious technique necessitates consistent sample preparation, data acquisition, and analyses by laboratory personnel in order to achieve accurate results on strain relatedness between samples. Despite its use as a gold standard for over two decades, the long turnaround time to results is a limitation that delays informing infection control teams for decision making at hospitals. Furthermore, due to the requirement of many reagents for sample preparation and data acquisition, it is not realistically applicable as a technique to repeatedly collect data from environmental and patient screening samples, to gain a bigger picture on prevalence of pathogens like VRE in the hospital setting. FTIR spectroscopy has the ability to shorten the turnaround time to results by acquiring spectral data and producing results within the same day that isolated colonies are obtained on agar plates. The lack of reagents required for spectral acquisition also eases the financial burden that is incurred in running more samples, thereby enabling more samples to be screened and analyzed than what is possible with PFGE. These two advantages that FTIR spectroscopy holds make it an appropriate high-throughput screening tool.

PFGE determines the degree of relatedness between two isolates based on the number of PFGE band differences, where a lower number of band differences between isolates indicates a higher degree of similarity between the isolates, although it is known that the strain type similarity of highly recombinant species may be poorly reflected by this strain type determination criterion. While some genetic mutations and gene transfers can greatly affect the structure or metabolic status of the microorganism, which may be observed as altered FTIR spectral features, other genetic alterations do not incur any phenotypic change, and thus FTIR spectra remain unchanged. The opposite can also occur, where phenotypic changes observable in the FTIR spectra may occur despite a lack of change in PFGE band patterns. Like the PFGE method, which relies on a set of criteria regarding the PFGE band patterns to determine whether isolates are clonal or not, the method that was developed in this thesis for strain typing of VRE using FTIR spectroscopy relied on cut-off values that were chosen, in analyses of the spectra of VRE outbreak related isolates that discriminated between two pulsotypes, AA and CC. The method utilized squared Mahalanobis distance between isolates as a measure of the spectral similarity between two isolates, where <500

squared distance indicated that the two isolates were spectrally indistinguishable. At the time this method was developed, no other genotypic data, such as MLST and whole genome SNP analyses, were available for consideration. Although this method and cut-off values should be further evaluated in relation to additional genotypic and possibly phenotypic characteristics, the retrospective and prospective studies presented in this thesis indicated that the determination of spectral similarity between isolates by transflection FTIR spectroscopy provides a reliable method for identifying isolates that are possibly implicated in an outbreak. This works on the basis that few to no genetic or phenotypic changes occur when transmission occurs in an outbreak, where the pathogen spreads quickly across the environment and among patients. As VRE adapts to its environment, changes in its phenotypic characteristics may occur, and hence repeated isolation and spectral characterization of samples from a carrier or infected patient may be useful in evaluating the stability of the spectral profile. Correlation between whole genome sequencing data and FTIR spectral data can also identify which gene mutations, acquisition or deletion through horizontal gene transfer have an effect on phenotypic characteristics.

The use of FTIR spectra for strain type characterization at the subspecies level, as well as for species identification, requires standardized sample preparation since spectral differences can result from differences in culturing conditions such as media composition, as well as time and temperature of incubation. Under standardized conditions, spectral differences observed between samples can be attributed to their inherent metabolic and structural differences. These biochemical differences which are reflected in the spectral fingerprint enable spectral discrimination among various microbial species, for their identification. One of the objectives of the research presented in this thesis was to evaluate whether FTIR spectra acquired in the transflection mode are satisfactory for purposes of microbial identification. While a much more extensive side-by-side comparison against transmission and ATR spectral acquisition modes would be required to fully evaluate the advantages and limitations of transflection FTIR spectroscopy for microbial identification applications, the spectral distortions often mentioned as a limitation of the transflection mode did not have a negative impact on its ability in microbial discrimination at species and subspecies level.

From a practical perspective, a key advantage of transflection FTIR spectroscopy is the capability to acquire spectra from microbial samples deposited on a wide variety of disposable or

reusable substrates, such as low-E glass and aluminum plates. Using a 96-well plate format, an automated transflection-based FTIR system could be produced similarly to the Bruker IR Biotyper, which is an automated transmission-based FTIR spectrometer that provides a reusable 96-well silicon plate for spectral acquisition. In contrast, automation of spectral acquisition in the ATR mode is harder to achieve.

FTIR spectroscopy observes changes in metabolism or gene expression when microbial isolates are cultured on media with different compositions or when culturing condition are changed. The FTIR spectra reflect the changes that the microorganisms utilize in order to adapt and survive, and therefore can be useful in detecting changes that are implicated in unfavorable conditions, such as the presence of antimicrobial agents. Additionally, growth of microorganisms on selective or differential media, which microbiologists employ to inhibit growth of certain organisms or observe differences among microorganisms based on their capability to metabolise specific compounds, may also be taken advantage of to aid FTIR spectroscopic discrimination among certain organisms. As an example, the use of MacConkey agar was found to enhance spectral differences between Gram-negative isolates, such as *E. coli* and *Shigella* species. This general principle laid the foundation for the novel FTIR spectroscopy-based MRSA identification method presented in this thesis. The addition of the antibiotic cefoxitin in blood agar plates resulted in MRSA identification with high sensitivity and specificity, by successfully differentiating MRSA from MSSA and CoNS based on the isolate growth and spectral differences respectively. The antibiotics colistin and nalidixic acid were also added to the culture plates to inhibit the growth of Gram-negative bacteria, for selective isolation of Gram-positive bacteria. The concept of acquiring FTIR spectra for identification of isolated colonies cultured on blood agar plates with antimicrobial agents may be extended to investigate the use of carbapenem or vancomycin in culture media for the identification of carbapenem-resistant Enterobacteriaceae (CRE) or VRE by FTIR spectroscopy.

In clinical labs, visual detection of MRSA and VRE isolates on chromogenic agar provides a quick confirmation of their presence in samples. However, when only relying on visual cues, without confirmation by techniques like MALDI-TOF MS or biochemical assays, false positive results may occur. While such instances were rarely encountered throughout years acquiring FTIR spectral data from isolates grown on chromogenic agar, a few violet colored isolates on

chromogenic agar were identified as non-*E. faecium* species by FTIR spectral analysis. MALDI-TOF MS later identified these isolates that were spectrally not *E. faecium* as *Enterococcus gallinarum* and *Lactobacillus* species, confirming the presence of non-target pathogens. FTIR spectral acquisition from well-isolated colonies can be useful as a one-step method that can provide both species identification and strain type characterization. Development of a FTIR spectral database with isolates cultured on chromogenic agar for species identification for the specific purpose of MRSA and VRE detection may allow for a smooth integration of FTIR spectroscopy for screening and characterization of organisms specifically for surveillance purposes. The use of chromogens that change color in the presence of specific enzymes possessed by target pathogens increases the cost of culture plates, relative to generic media, such as Columbia agar with sheep blood or Mueller Hinton agar. By acquiring FTIR spectra from isolates cultured on antibiotic containing media, the need for chromogens to visually differentiate between target pathogens such as MRSA from MRS-CoNS, and *E. faecium* from *E. faecalis* becomes unnecessary, thereby reducing the expense incurred for chromogenic agar plates. Furthermore, as mentioned before, the spectra can be used for both identification and strain type characterization, given that the culturing conditions used are standardized and consistent between samples. Cefoxitin presence in the agar can also provide a presumptive result for resistant strains, which can later be confirmed with antimicrobial susceptibility tests for a panel of relevant agents specific to MRSA. Growth in the presence of cefoxitin is used to determine whether isolates are resistant against all β-lactam antimicrobial agents for *S. aureus*, and thus used for MRSA detection. The addition of cefoxitin in blood agar resulted in subtle changes in the FTIR spectra of MRSA, which can be used to study the change related to activation of antimicrobial resistance genes such as *mecA*. Since FTIR spectroscopy provides a rapid and reliable way for microorganisms to be characterized, studies involving spectral analyses on the effects of antimicrobial agents at sub-inhibitory levels may be useful in understanding the metabolic differences between sensitive and resistant strains, which may be correlated to gene expressions or activation.

Metabolic and structural changes that occur under environmental stress or result from mutation in the genome may be observed by FTIR spectroscopy and correlated to observed genotypic alterations. Spectroscopic data and genotypic data acquired on VRE pulsotypes confirmed there were immense differences in their biochemical makeup and the genome itself. They also proved the sensitivity of both FTIR spectroscopy and ^{1}H and ^{31}P HR-MAS NMR spectroscopy to observe

differences that discriminate organisms at the subspecies level. FTIR spectroscopy can become useful in microbiology laboratories as a high-throughput screening technology, to identify isolates and spectrally compare them to other samples for relatedness, in the context of outbreak detection and surveillance, for both clinical and food-borne pathogens. Detection of isolates that are spectrally indistinguishable, is just as important as identifying isolates with unique spectral features; in both cases, selected isolates can be followed up by more laborious phenotypic and genotypic methods for further characterization.

As a final comment, classification based on spectral features does not necessarily follow the current taxonomic classification system, which is based on the combination of phylogenetic, phenotypic and genotypic characteristics. The microbial taxonomy is ever evolving, and involves constant (re-)classification, as more information on microbial organisms is collected. As part of this process, FTIR spectral characteristics may provide valuable data in characterizing and classifying bacterial species.

Chapter 8. Conclusion

The objective of the thesis was to evaluate the use of transflection FTIR spectroscopy for microbial applications, specifically for identification and discrimination at the species and subspecies levels. The results presented in the thesis demonstrated that like other spectral acquisition modes (i.e. transmission and ATR) that have been widely employed to acquire FTIR spectra of intact microbial cells, the transflection mode provides spectra of sufficient quality to serve as "whole-organism" fingerprints of microorganisms, for discrimination at the species and strain-type level. A transflection FTIR spectral database and a multi-tired spectral classification were developed for microbial identification to the species level. Correct identification of > 98 % of *E. faecalis, E. faecium* and *S. aureus* isolates (1103 isolates combined between the three species) demonstrated that species identification was possible upon sufficient spectral representation of the species in the spectral database. While identification with high concordance relative to reference methods like MALDI-TOF MS and VITEK 2 was achieved for these isolates, not all genera and species could be identified with the same level of concordance. This was particularly the case for Gram-negative genera within the Enterobacteriaceae family and for species within the *Staphylococcus* genus, with the exception of *S. aureus*, and may be attributed to a number of factors, including low representation in the spectral database, spectral similarity between certain poorly identified genera, as well as the diversity in genotypic and phenotypic characteristics within species. Although extensive FTIR spectral analyses and understanding of the characteristics of these genera and species are required to fully evaluate whether successful spectral discrimination and identification can be achieved, improvements in identification results were observed with increased spectral representation in the database, as previously reported by others. The addition of antibiotics such as cefoxitin, colistin and nalidixic acid into the Columbia blood agar growth medium did not affect the performance achieved for the identification of *S. aureus* using the transflection FTIR spectral database enabling accurate MRSA identification through the combined use of antibiotics in the culture medium and transflection FTIR spectroscopy for discrimination of *S. aureus* from CoNS (98 % categorical agreement). Furthermore at the subspecies level, development of a strain typing method using transflection FTIR spectroscopy in conjunction with multivariate statistical analyses revealed its capability and usefulness as a rapid screening technique for identifying isolates that may be part of an outbreak. In a 6-month prospective study, VRE isolates obtained from patient and environmental screening samples were spectrally analyzed

on a weekly basis and isolates classified as spectrally indistinguishable were later confirmed by PFGE analysis. Proton and ^{31}P NMR spectra acquired from microbial isolates by HR-MAS NMR spectroscopy demonstrated that discrimination at the species and sub-species levels was achievable, similarly to FTIR spectral discrimination. Owing to the higher chemical specificity of NMR spectroscopy by comparison to vibrational spectroscopy, potential biomarkers that contributed to successful spectral discrimination between two VRE pulsotypes by both NMR and FTIR spectroscopy were tentatively identified in the averaged ^{1}H and averaged ^{31}P NMR spectra of VRE isolates belonging to each of these pulsotypes, however, analyses by multidientional NMR experiments and mass spectrometer would be required for definitive identification of biomarkers. Whole genome sequencing data indicated that there are substantial genetic differences between the VRE pulsotypes. However, at this time, the spectroscopic differences identified could not be correlated to specific genomic differences between the VRE pulsotypes. The preliminary work conducted in an attempt to correlate between spectroscopic and genotypic data demonstrates how the spectral discriminatory capabilities and observations made at the sub-species level using FTIR spectroscopy provides a rapid and consistent method for characterizing microbial isolates in the live state, which can be supplemented by various phenotypic and genotypic techniques.

Overall, the results from the series of experiments presented in this thesis demonstrated the capabilities of using transflection FTIR spectroscopy as a rapid and routine technique for microbial identification and strain typing applications, in particular for determination of strain relatedness among isolates for outbreak surveillance and epidemiological purposes. Future work should be directed toward 1. selective, and comprehensive spectral analyses on Gram-negative genera and staphylococcal species by transflection-FTIR spectroscopy for evaluation of its capabilities in their identification, 2. multi-centre evaluation of the transflection FTIR spectroscopy-based microbial identification method, 3. application of the protocol employed for MRSA detection using antibiotic-selective agar and FTIR spectroscopy, for detection of other antibiotic resistant microorganisms, such as vancomycin-resistant enterococci, and extended-spectrum β-lactamase-positive Gram-negative microorganisms, 4. continuation of biomarker elucidation research, focusing on cell wall constituents, using multidimensional NMR spectroscopy, mass spectrometry and whole genome sequencing, 5. combined use of whole-organism fingerprinting techniques (FTIR spectroscopy and HR-MAS NMR spectroscopy) for dynamic studies, such as observing the interaction and effect of antibiotics and potential drug candidates on live bacteria and yeast cell.

APPENDIX

Supplementary Table 1 Update identification results after the addition of isolates into training set for GP genera

	Training							Validation+ TEST						
	TOTAL	CORRECT (n, %)		NO ID (n, %)		MIS ID (n, %)		TOTAL	CORRECT (n, %)		NO ID (n, %)		MIS ID (n, %)	
Aerococcus								2	0	0	1	50	1	50
Bacillus	15	15	100	0		0		20	10	50	0	0	10	50
Corynebacterium	9	9	100	0		0		9	8	88.9	1	11.1	0	0
Dermabacter								3	0	0	0	0	3	100
Enterococcus	89	89	100	0		0		634	633	99.8	0	0	1	0.2
Gordonia								1	0	0	0	0	1	100
Kocuria								3	0	0	1	33.3	2	66.7
Kytococcus								1	0	0	1	100	0	0
Lactobacillus								4	0	0	1	25	3	75
Lactococcus								1	0	0	0	0	1	100
Listeria	31	31	100	0		0		26	25	96.2	0	0	1	3.8
Micrococcus	5	5	100	0		0		14	13	92.9	1	7.1	0	0
Rhodococcus								1	0	0	1	100	0	0
Rothia								1	0	0	0	0	1	100
Staphylococcus	332	332	100	0		0		1276	1265	99.1	0	0	11	0.9
Streptococcus	63	63	100	0		0		92	85	92.4	0	0	7	7.6
Turicella								2	0	0	0	0	2	100
Total	**544**	**544**	**100**	**0**	**0**	**0**	**0**	**2090**	**2039**	**97.6**	**7**	**0.3**	**44**	**2.1**
Represented Genera	**544**	**544**	**100**	**0**	**0**	**0**	**0**	**2071**	**2039**	**98.5**	**2**	**0.1**	**30**	**1.4**
Other genera	0	0		0		0		19	5	**26.3**		**0**	14	**73.7**

Supplementary Table 2 Updated identification results after addition of isolates into training set for GN genera

	Training							Validation+TEST						
	Total		Correct (n, %)		No ID (n, %)	Mis ID (n, %)		Total		Correct (n, %)		No ID (n, %)	Mis ID (n, %)	
Achromobacter	22	22	100	0	0	0	0	19	18	94.7	1	5.3	0	0
Acinetobacter	11	11	100	0	0	0	0	11	6	54.5	0	0	5	45.5
Aeromonas	19	15	78.9	0	0	4	21.1	8	2	25	0	0	6	75
Brevundimonas								2	0	0	0	0	2	100
Burkholderia	20	18	90	2	10	0	0	4	1	25	2	50	1	25
Chryseobacterium								9	0	0	3	33.3	6	66.7
Citrobacter	26	18	69.2	8	30.8	0	0	7	1	14.3	3	42.9	3	42.9
Cupriavidus								2	0	0	1	50	1	50
Delftia								1	0	0	0	0	1	100
Edwardsiella								1	0	0	1	100	0	0
Enterobacter	30	24	80	5	16.7	1	3.3	30	14	46.7	12	40	4	13.3
Escherichia	59	54	91.5	5	8.5	0	0	91	58	63.7	22	24.2	11	12.1
Kingella								3	0	0	0	0	3	100
Klebsiella	48	47	97.9	1	2.1	0	0	36	28	77.8	4	11.1	4	11.1
Moraxella								1	0	0	1	100	0	0
Morganella								12	0	0	5	41.7	7	58.3
Ochrobactrum								7	0	0	4	57.1	3	42.9
Pantoea								10	0	0	4	40	6	60
Paracoccus								2	0	0	1	50	1	50
Pasteurella								1	0	0	0	0	1	100
Plesiomonas								4	0	0	4	100	0	0
Prevotella								1	0	0	0	0	1	100
Proteus	11	10	90.9		0	1	9.1	9	6	66.7	1	11.1	2	22.2
Providencia								3	0	0	3	100	0	0
Pseudomonas	40	35	87.5	1	2.5	4	10	56	41	73.2	5	8.9	10	17.9
Raoultella								4	0	0	3	75	1	25
Salmonella	60	50	83.3	10	16.7	0	0	110	83	75.5	21	19.1	6	5.5
Serratia	0	0						16	0	0	7	43.8	9	56.3
Shigella	30	29	96.7	1	3.3	0	0	21	13	61.9	6	28.6	2	9.5
Sphingomonas								2	0	0	0	0	2	100
Stenotrophomonas	30	27	90	1	3.3	2	6.7	16	10	62.5	1	6.3	5	31.3
Vibrio								8	0	0	5	62.5	3	37.5
Yersinia								4	0	0	2	50	2	50
TOTAL	406	360	88.7	34	8.4	12	3	511	281	55	122	23.9	108	21.1
Represented genera (w/o *Serratia*)	406	360	88.7	34	8.4	12	3	418	281	67.2	78	18.7	59	14.1
non-represented GN genera								93	0	0	44	47.3	49	52.7

Supplementary Table 3 Identification results after update of spectral database with addition of isolates for CoNS subgroups

CoNS Cluster groups	**Training**							**Validation+TEST**						
	TOTAL	CORRECT (n, %)		No ID (n, %)		MIS ID (n, %)		TOTAL	CORRECT (n, %)		No ID (n, %)		MIS ID (n, %)	
cohnii group	11	10	90.9	1	9.1	0	0	8	6	75	2	25	0	0
epidermidis group	70	68	97.1	2	2.9	0	0	111	97	87.4	11	9.9	3	2.7
haemolyticus group	50	47	94	0	0	3	6	43	38	88.4	4	9.3	1	2.3
lugdunensis group	10	10	100	0	0	0	0	11	7	63.6	3	27.3	1	9.1
saprophyticus group	10	8	80	2	20	0	0	9	3	33.3	5	55.6	1	11.1
warneri group	15	11	73.3	4	26.7	0	0	12	2	16.7	5	41.7	5	41.7
Total	**155**	**144**	**92.9**	**8**	**5.2**	**3**	**1.9**	**186**	**147**	**79**	**28**	**15.1**	**11**	**5.9**

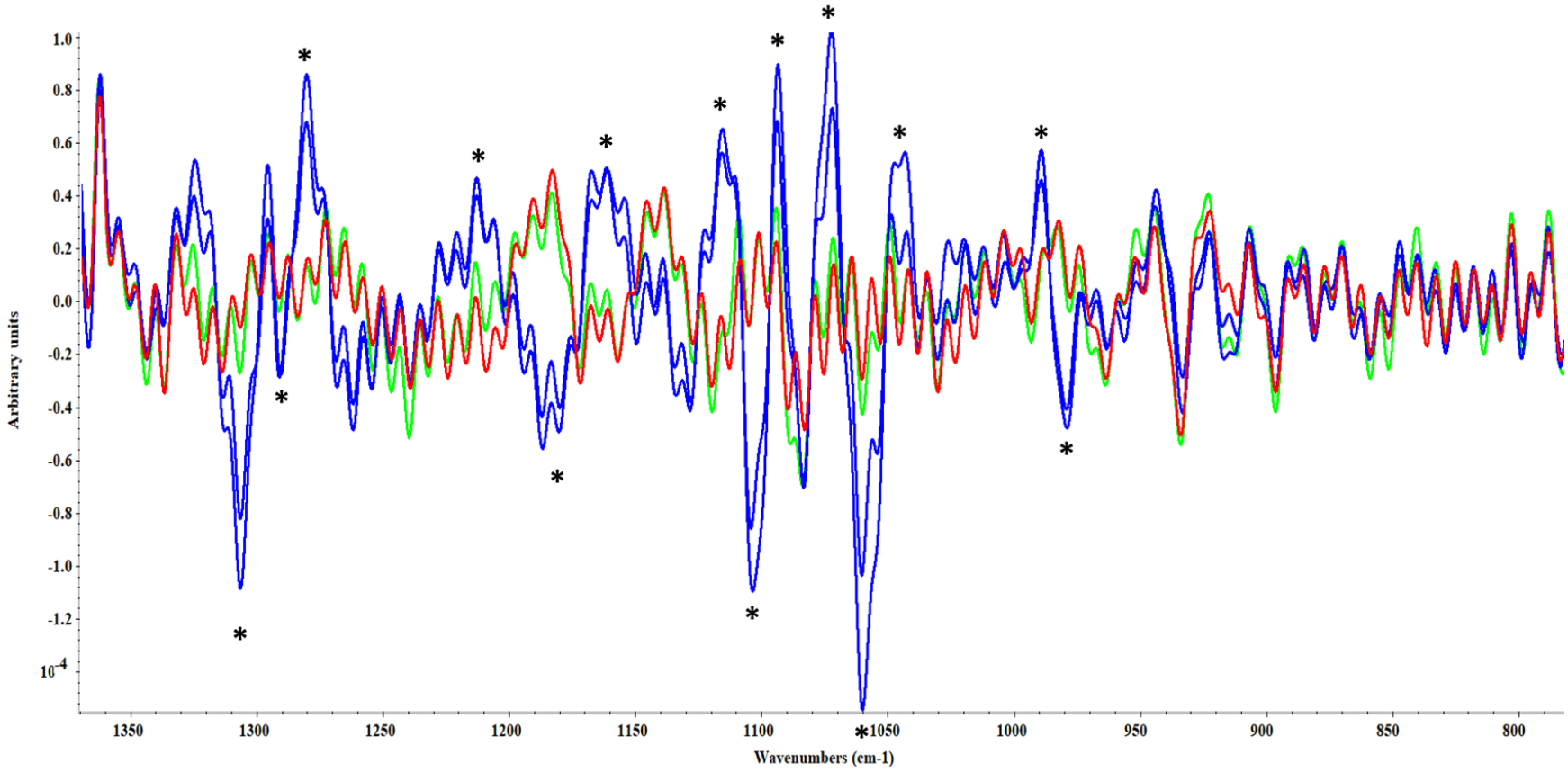
Supplementary Table 4 Species identification results after spectral database update

	Training						Validation+TEST					
	Total	Correct (n, %)		No ID (n, %)		Mis ID (n, %)		Total	Correct (n, %)		No ID (n, %)		Mis ID (n, %)	
E. avium								1	0	0	0	0	1	100
E. casseliflavis								3	0	0	0	0	3	100
E. faecalis	40	40	100	0		0		85	85	100	0	0	0	0
E. faecium	40	40	100	0		0		519	513	98.8	0	0	6	1.2
E.gallinarum								18	0	0	0	0	18	100
L. grayi	2	0	0	2	100	0	0	1	0	0	1	100	0	0
L. monocytogenes	24	24	100	0		0	0	20	18	90	1	5	1	5
L.ivanovii								1	0	0	1	100	0	0
L. seeligeri	1	0		1	100	0	0	0	0		0		0	
L. welshimeri	2	0		2	100	0	0	2	0	0	0	0	2	100
L. innocua	2	0		2	100	0	0	2	0	0	2	100	0	0
S. aureus	166	166	100	0		0	0	1022	1005	98.3	0	0	17	1.7
S. capitis	30	27	90	1	3.3	2	6.7	22	20	90.9	0	0	2	9.1
S. caprae								8	0	0	6	75	2	25
S. carnosus								1	0	0	1	100	0	0
S. caseolyticus								1	0	0	0	0	1	100
S. cohnii	11	11	100	0		0		8	6	75	2	25	0	0
S. epidermidis	40	39	97.5	1	2.5	0		107	96	89.7	6	5.6	5	4.7
S. equorum	0					0		1	0	0	1	100	0	0
S. haemolyticus	20	18	90	0		2	10	21	15	71.4	2	9.5	4	19
S. hominis	30	29	96.7	0		1	3.3	45	42	93.3	2	4.4	1	2.2
S. lugdunensis	10	10	100	0	0	0		11	7	63.6	3	27.3	1	9.1
S. pasteuri								4	0	0	2	50	2	50
S. pseudintermedius								1	0	0	0	0	1	100
S. saprophyticus	10	8	80	2	20	0		6	3	50	2	33.3	1	16.7
S. simulans								8	0	0	5	62.5	3	37.5
S. vitulinus								1	0	0	0	0	1	100
S. warneri	15	11	73.3	4	26.7	0	0	11	4	36.4	3	27.3	4	36.4
S. xylosus	0					0		2	0	0	1	50	1	50
S. boydii								1	0	0	0	0	1	100
S. dysenteriae								2	0	0	0	0	2	100
S. flexneri	18	17	94.4	0	0	1	5.6	7	6	85.7	0	0	1	14.3
S. sonnei	12	12	100	0	0	0	0	10	7	70	0	0	3	30
Total	473	452	95.6	15	3.2	6	1.3	1952	1827	93.6	41	2.1	84	4.3

Supplementary Table 5 List of isolates identified as outliers by transflection-FTIR spectral database indicating potential sample mix-up among isolates obtained from same sample source

Sample ID	**Original ID**	**Updated Database**			**Original Database**		
		Predicted	Probability	Sample Type	Predicted	Probability	Sample Type
XVIII291-2	*Klebsiella pneumoniae*	*Acinetobacter*	97.5	Test	*Acinetobacter*	93.9	Test
XVIII 291-3	*Acinetobacter baumanii complex*	*Klebsiella*	99.9	Test	*Klebsiella*	99.9	Test
TVII779-2	*Enterobacter cloacae-complex*	*Stenotrophomonas*	75.9	Training	GN - no further id	99.9	Training
TVII779-4	*Pseudomonas aeruginosa*	*Enterobacter*	99.9	Validation	*Enterobacter*	99.9	Test
TVIIII112-1	*Enterobacter cloacae-complex*	*Staphylococcus aureus*	95.4	Test	*Staphylococcus aureus*	94.7	Test
TVIIII112-2	*Staphylococcus aureus*	Enterobacteriaceae	99.9	Test	Enterobacteriaceae	86.3	Test

Supplementart Figure 1 Overlay of second derivative spectra of averaged Bacillus cereus-group (blue), Bacillus megaterium (blue), Staphylococcus aureus (green) in relation to Bacillus isolate outlier (red) in spectral range 1350-800 cm^{-1}



Supplementary Table 6 Formulation of Columbia agar with 5 % Sheep Blood

Ingredients	**Approximate Formula /Litre**
Agar	10.0 g
Sodium Chloride	5.0 g
Special Peptone	23.0 g
Starch	1.0 g
Sheep Blood	50 ml

Formula of Columbia agar with 5 % sheep blood from Oxoid.

Supplementary Table 7 Table of results for Staphylococcus isolates cultured on modified, antibiotic containing media, identified as S. aureus or CoNS using transflection FTIR spectral database

		BAP			**CNA_4FOX**			**CNA_8FOX**		
Species	Sensitivity	*n*	Correct	(%)	*n*	Correct	(%)	*n*	Correct	(%)
S. aureus	MRSA	56	56	100	56	56	100	51	51	100
S. aureus	MSSA	43	43	100	2	2	100			
CoNS										
S. capitis	MRS	5	5	100	5	5	100	5	5	100
S. capitis	MSS	19	19	100						
S. caprae	MSS	2	2	100						
S. cohnii	MRS	1	1	100	1	1	100	1	1	100
S. epidermidis	MRS	27	27	100	27	24	88.9	27	21	77.8
S. epidermidis	MSS	22	22	100						
S. haemolyticus	MRS	9	9	100	9	9	100	8	8	100
S. haemolyticus	MSS	3	3	100						
S. hominis	MRS	10	10	100	7	7	100	4	4	100
S. hominis	MSS	15	15	100	1	1	100			
S. lugdunensis	MRS	3	3	100	1	1	100			
S. lugdunensis	MSS	7	7	100	1	1	100			
S. saprophyticus	MSS									
S. species	MRS	1	1	100	1	1	100	3	3	100
S. species	MSS	2	2	100						
S. warneri	MSS	4	4	100						
Total		229	229	100	111	108	97.3	99	93	93.9

All staphylococcal species achieved 100 % correct identification as CoNS on all media types with the exception of *S.epidermidis*, which achieved 88.9 % and 77.8 % on CNA_4FOX and CNA_8FOX respectively. Empty cells indicate no growth, and therefore no numbers to report on correct identification as *S. aureus* or CoNS.

Supplementary Table 8 Isolate and whole genome sequencing information

SEQID	Curator Flag	rMLST Result	MASH Reference Genome	MASH Num Matching Hashes	16S_result	N50	Num Contigs	Total Length	Average Coverage Depth	Gene Seekr Profile
SEQ-161	PASS	new	*Enterococcus sp.*	802/1000	*Enterococcus faecium*	6,878	708	2619057	65.77	IGS
SEQ-162	PASS	new	*Enterococcus sp.*	795/1000	*Enterococcus faecium*	3,229	1062	2394399	52.16	IGS
SEQ-163	PASS	new	*Enterococcus faecium*	778/1000	*Enterococcus faecium*	3,266	1035	2319444	92.26	IGS
SEQ-164	PASS	new	*Enterococcus sp.*	794/1000	*Enterococcus faecium*	17,854	306	2750870	118.56	IGS
SEQ-165	FAIL	new	*Enterococcus faecium*	744/1000	*Enterococcus faecium*	1,252	1339	1523822	21.18	IGS
SEQ-166	FAIL	new	*Enterococcus faecium*	741/1000	*Enterococcus faecium*	1,484	1338	1723183	15.47	IGS
SEQ-167	PASS	new	*Enterococcus faecium*	806/1000	*Enterococcus faecium*	4,046	993	2473755	43.59	IGS
SEQ-169	PASS	new	*Enterococcus faecium*	652/1000	*Enterococcus faecium*	12,612	410	2637120	179.81	IGS
SEQ-170	PASS	new	*Enterococcus faecium*	660/1000	*Enterococcus faecium*	6,907	668	2547837	129.14	IGS
SEQ-171	PASS	new	*Enterococcus faecium*	728/1000	*Enterococcus faecium*	9,066	536	2655477	78.29	IGS
SEQ-172	PASS	18571	*Enterococcus faecium*	633/1000	*Enterococcus faecium*	21,637	315	2754000	159.25	IGS
SEQ-173	PASS	18440	*Enterococcus faecium*	703/1000	*Enterococcus faecium*	14,555	449	2814927	114.97	IGS
SEQ-174	PASS	18440	*Enterococcus faecium*	777/1000	*Enterococcus faecium*	23,231	349	2864784	89.94	IGS
SEQ-175	PASS	18440	*Enterococcus faecium*	715/1000	*Enterococcus faecium*	16,178	436	2849384	128.82	IGS
SEQ-176	PASS	new	*Enterococcus faecium*	749/1000	*Enterococcus faecium*	8,068	669	2776266	69.75	IGS
SEQ-177	PASS	new	*Enterococcus faecium*	727/1000	*Enterococcus faecium*	7,670	680	2748111	71.68	IGS
SEQ-178	PASS	18440	*Enterococcus faecium*	650/1000	*Enterococcus faecium*	12,758	485	2820291	191.31	IGS
SEQ-179	PASS	18440	*Enterococcus faecium*	648/1000	*Enterococcus faecium*	5,794	802	2675592	107.34	IGS
SEQ-180	PASS	18440	*Enterococcus faecium*	778/1000	*Enterococcus faecium*	15,758	424	2829692	68.43	IGS
SEQ-181	PASS	18440	*Enterococcus faecium*	613/1000	*Enterococcus faecium*	18,828	365	2861568	260.05	IGS
SEQ-182	PASS	new	*Enterococcus faecium*	755/1000	*Enterococcus faecium*	4,109	1025	2595528	70.25	IGS
SEQ-183	FAIL	new	*Enterococcus faecium*	723/1000	*Enterococcus faecium*	1,085	1339	1388195	26.96	IGS
SEQ-184	PASS	new	*Enterococcus faecium*	731/1000	*Enterococcus faecium*	3,319	1156	2531396	78.42	IGS
SEQ-190	PASS	18425	*Enterococcus faecium*	797/1000	*Enterococcus faecium*	51,756	144	2790758	106.29	IGS

(continued)

SEQID	Curator Flag	rMLST Result	MASH Reference Genome	MASH Num Matching Hashes	16S_result	N50	Num Contigs	Total Length	Average Coverage Depth	Gene Seekr Profile
SEQ-185	PASS	18567	*Enterococcus faecalis*	772/1000	*Enterococcus faecalis*	44,749	217	3278265	201.02	IGS
SEQ-186	CONT.	18687	*Enterococcus faecalis*	774/1000	*Enterococcus faecalis*	15,746	339	2869070	76.15	IGS
SEQ-188	PASS	new	*Enterococcus sp*	868/1000	*Enterococcus faecium*	74,933	91	2736655	153.23	IGS
SEQ-189	PASS	new	*Enterococcus faecalis*	729/1000	*Enterococcus faecium*	95,299	60	2816242	217.82	IGS

Supplementary Table 9 Antimicrobial resistance genes identified from whole genome sequencing

Strain	Gene	Allele	Resistance	% Identity	Contig	Location	Plasmid Incompatibility Sets
SEQ-0161	*aac(6')*	1	aminoglycoside	99.64	Contig_458_17.8837	chromosome	ND
	ant(6)-Ia	1	aminoglycoside	100	Contig_791_55.3945	1642	ND;rep_cluster_889
	dfrG	1	trimethoprim	100	Contig_661_22.9959	chromosome	ND
	erm(B)	9	macrolide	100	Contig_791_55.3945	1642	ND;rep_cluster_889
	msr(C)	1	macrolide	98.99	Contig_59_34.0246	chromosome	ND
	tet(L)	2	tetracycline	97.46	Contig_663_12.668	2351	rep_cluster_1018
	tet(M)	10	tetracycline	75.94	Contig_690_53.2235	1866	ND
	VanHAX	2	glycopeptide	99.92	Contig_364_75.2906	1656	ND;rep_cluster_943
SEQ-0162	*ant(6)-Ia*	1	aminoglycoside	100	Contig_1246_47.4491	1642	ND;rep_cluster_889
	dfrG	1	trimethoprim	100	Contig_1383_7.47338	chromosome	ND
	erm(B)	9	macrolide	100	Contig_1246_47.4491	1642	ND;rep_cluster_889
	msr(C)	1	macrolide	98.99	Contig_260_19.6136	chromosome	ND
	VanHAX	2	glycopeptide	99.92	Contig_704_68.7772	1656	ND;rep_cluster_943
SEQ-0163	*aac(6')*	1	aminoglycoside	84.34	Contig_832_10.0909	chromosome	ND
	ant(6)-Ia	1	aminoglycoside	100	Contig_1336_39.2574	1642	ND;rep_cluster_889
	erm(B)	9	macrolide	100	Contig_1336_39.2574	1642	ND;rep_cluster_889
	msr(C)	1	macrolide	98.99	Contig_861_15.9859	chromosome	ND
	VanHAX	2	glycopeptide	99.92	Contig_262_79.0954	1656	ND;rep_cluster_943
SEQ-0164	*aac(6')*	1	aminoglycoside	99.64	Contig_158_67.4439	chromosome	ND
	ant(6)-Ia	1	aminoglycoside	100	Contig_359_167.969	1642	ND;rep_cluster_889
	dfrG	1	trimethoprim	100	Contig_295_75.4794	chromosome	ND
	erm(B)	9	macrolide	100	Contig_359_167.969	1642	ND;rep_cluster_889
	msr(C)	1	macrolide	98.99	Contig_298_91.6916	chromosome	ND
	tet(L)	2	tetracycline	97.46	Contig_98_55.4671	2351	ND;rep_cluster_1018
	VanHAX	2	glycopeptide	99.92	Contig_314_180.669	1656	ND

(continued)

Strain	Gene	Allele	Resistance	% Identity	Contig	Location	Plasmid Incompatibility Sets
SEQ-0167	*aac(6')*	1	aminoglycoside	99.64	Contig_1257_19.1972	chromosome	ND
	ant(6)-Ia	1	aminoglycoside	100	Contig_1349_48.6125	1642	ND;rep_cluster_889
	dfrG	1	trimethoprim	100	Contig_878_18.4869	chromosome	ND
	erm(B)	9	macrolide	100	Contig_1349_48.6125	1642	ND;rep_cluster_889
	msr(C)	1	macrolide	98.31	Contig_743_19.1642	chromosome	ND
	tet(L)	2	tetracycline	97.46	Contig_961_13.004	2351	rep_cluster_1018
	VanHAX	2	glycopeptide	99.92	Contig_1326_53.9697	1656	ND;rep_cluster_943
SEQ-0169	*aac(6')*	1	aminoglycoside	99.64	Contig_303_55.1766	chromosome	ND
	ant(6)-Ia	1	aminoglycoside	100	Contig_472_149.889	1642	ND;rep_cluster_889
	dfrG	1	trimethoprim	100	Contig_381_52.9732	chromosome	ND
	erm(B)	9	macrolide	100	Contig_472_149.889	1642	ND;rep_cluster_889
	msr(C)	1	macrolide	98.99	Contig_390_74.1744	chromosome	ND
	tet(L)	2	tetracycline	97.46	Contig_295_36.786	2351	rep_cluster_1018
	VanHAX	2	glycopeptide	99.92	Contig_191_264.76	1656	ND;rep_cluster_943
SEQ-0170	*aac(6')*	1	aminoglycoside	99.64	Contig_493_34.3291	chromosome	ND
	ant(6)-Ia	1	aminoglycoside	100	Contig_688_64.7384	1642	ND;rep_cluster_889
	dfrG	1	trimethoprim	100	Contig_640_30.8107	chromosome	ND
	erm(B)	9	macrolide	100	Contig_688_64.7384	1642	ND;rep_cluster_889
	msr(C)	1	macrolide	98.99	Contig_686_46.8572	chromosome	ND
	tet(L)	2	tetracycline	89.32	Contig_150_29.9716	1866	ND
	VanHAX	2	glycopeptide	99.92	Contig_139_97.5167	1656	ND;rep_cluster_943
SEQ-0171	*aac(6')*	1	aminoglycoside	99.64	Contig_511_45.9406	chromosome	ND
	ant(6)-Ia	1	aminoglycoside	100	Contig_206_62.2681	1649	ND;rep_cluster_893
	aph(3')-III	1	aminoglycoside	100	Contig_206_62.2681	1649	ND;rep_cluster_893
	dfrG	1	trimethoprim	100	Contig_487_40.8316	chromosome	ND
	erm(B)	9	macrolide	100	Contig_567_34.3453	1649	ND;rep_cluster_893
	msr(C)	1	macrolide	98.99	Contig_585_62.5015	chromosome	ND
	tet(L)	2	tetracycline	97.46	Contig_197_27.9486	2351	ND;rep_cluster_1018
	VanHAX	2	glycopeptide	99.92	Contig_426_77.9289	1656	ND;rep_cluster_889

(continued)

Strain	Gene	Allele	Resistance	% Identity	Contig	Location	Plasmid Incompatibility Sets
SEQ-0172	*aac(6')*	1	aminoglycoside	99.64	Contig_327_78.422	chromosome	ND
	ant(6)-Ia	1	aminoglycoside	100	Contig_120_127.542	1649	ND;rep_cluster_889;rep_cluster_893
	aph(3')-III	1	aminoglycoside	100	Contig_120_127.542	1649	ND;rep_cluster_889;rep_cluster_893
	dfrG	1	trimethoprim	100	Contig_264_97.1105	chromosome	ND
	erm(B)	9	macrolide	100	Contig_280_89.5213	1649	ND;rep_cluster_889;rep_cluster_893
	msr(C)	1	macrolide	98.99	Contig_132_117.721	chromosome	ND
	tet(L)	2	tetracycline	97.46	Contig_278_51.6496	2351	ND
	VanHAX	2	glycopeptide	99.92	Contig_226_241.185	1656	ND;rep_cluster_889;rep_cluster_943
SEQ-0173	*aac(6')*	1	aminoglycoside	99.64	Contig_51_102.943	chromosome	ND
	ant(6)-Ia	1	aminoglycoside	81.08	Contig_158_139.834	chromosome	ND
	aph(3')-III	1	aminoglycoside	99.87	Contig_158_139.834	chromosome	ND
	dfrG	1	trimethoprim	84.94	Contig_485_16.3974	chromosome	ND
	erm(B)	9	macrolide	100	Contig_475_102.53	chromosome	ND
	msr(C)	1	macrolide	98.99	Contig_26_82.0792	chromosome	ND
	tet(L)	2	tetracycline	100	Contig_412_28.766	2351	ND
	tet(M)	10	tetracycline	96.46	Contig_96_96.7941	1866	ND;rep_cluster_185
	VanHAX	2	glycopeptide	99.85	Contig_348_237.365	novel_1	ND;rep_cluster_889
SEQ-0174	*aac(6')*	1	aminoglycoside	99.64	Contig_147_60.0504	chromosome	ND
	ant(6)-Ia	1	aminoglycoside	100	Contig_293_92.0368	3050	ND
	aph(3')-III	1	aminoglycoside	99.87	Contig_293_92.0368	3050	ND
	dfrG	1	trimethoprim	100	Contig_110_33.3913	chromosome	ND
	msr(C)	1	macrolide	98.99	Contig_318_79.0932	chromosome	ND
	tet(L)	2	tetracycline	100	Contig_422_30.3259	2351	ND
	tet(M)	10	tetracycline	96.46	Contig_104_90.991	1866	ND;rep_cluster_185
	VanHAX	2	glycopeptide	99.85	Contig_315_195.924	1656	ND;rep_cluster_889

(continued)

Strain	Gene	Allele	Resistance	% Identity	Contig	Location	Plasmid Incompatibility Sets
SEQ-0175	*aac(6')*	1	aminoglycoside	99.64	Contig_470_65.058	chromosome	ND
	ant(6)-Ia	1	aminoglycoside	100	Contig_110_125.419	3050	ND
	aph(3')-III	1	aminoglycoside	99.87	Contig_110_125.419	3050	ND
	cat(pC221)	1	phenicol	95.06	Contig_299_66.8309	chromosome	ND
	dfrG	1	trimethoprim	100	Contig_94_29.1689	chromosome	ND
	msr(C)	1	macrolide	98.99	Contig_32_64.5487	chromosome	ND
	tet(L)	2	tetracycline	100	Contig_504_35.0809	1866	ND;rep_cluster_185
	tet(M)	10	tetracycline	96.46	Contig_354_90.6888	1866	ND;rep_cluster_185
	VanHAX	2	glycopeptide	99.85	Contig_142_309.565	1649	ND;rep_cluster_889;rep_cluster_893
SEQ-0176	*aac(6')*	1	aminoglycoside	99.64	Contig_527_21.2307	chromosome	ND
	ant(6)-Ia	1	aminoglycoside	100	Contig_691_44.1414	3050	ND
	aph(3')-III	1	aminoglycoside	99.87	Contig_691_44.1414	3050	ND
	cat(pC221)	1	phenicol	97.69	Contig_535_96.6345	novel_0	ND;rep_cluster_1118
	dfrG	1	trimethoprim	100	Contig_603_11.034	chromosome	ND
	msr(C)	1	macrolide	98.99	Contig_44_28.2756	chromosome	ND
	tet(L)	2	tetracycline	100	Contig_818_20.3888	2351	ND
	tet(M)	10	tetracycline	96.46	Contig_173_31.7586	1866	ND;rep_cluster_185
	VanHAX	2	glycopeptide	99.85	Contig_148_211.69	1656	ND;rep_cluster_889
SEQ-0177	*aac(6')*	1	aminoglycoside	99.64	Contig_783_67.5223	chromosome	ND
	ant(6)-Ia	1	aminoglycoside	100	Contig_636_49.3069	3050	ND
	aph(3')-III	1	aminoglycoside	99.87	Contig_636_49.3069	3050	ND
	cat(pC221)	1	phenicol	93.21	Contig_335_37.3832	2351	ND
	erm(B)	9	macrolide	100	Contig_623_66.91	3050	ND
	msr(C)	1	macrolide	98.99	Contig_606_40.6414	chromosome	ND
	tet(M)	10	tetracycline	95.47	Contig_147_59.9389	1866	ND;rep_cluster_185
	VanHAX	2	glycopeptide	99.85	Contig_220_193.495	1649	ND;rep_cluster_889

(continued)

Strain	Gene	Allele	Resistance	% Identity	Contig	Location	Plasmid Incompatibility Sets
SEQ-0178	*aac(6')*	1	aminoglycoside	99.64	Contig_52_58.2458	chromosome	ND
	ant(6)-Ia	1	aminoglycoside	100	Contig_395_138.768	3050	ND
	aph(3')-III	1	aminoglycoside	99.87	Contig_395_138.768	3050	ND
	cat(pC221)	1	phenicol	97.69	Contig_382_239.889	novel_0	rep_cluster_1118
	dfrG	1	trimethoprim	100	Contig_133_27.4537	chromosome	ND
	msr(C)	1	macrolide	98.99	Contig_41_76.2842	chromosome	ND
	tet(M)	10	tetracycline	96.15	Contig_286_61.6505	chromosome	ND
	VanHAX	2	glycopeptide	99.85	Contig_127_506.993	1649	ND;rep_cluster_889
SEQ-0179	*aac(6')*	1	aminoglycoside	99.64	Contig_603_25.1776	chromosome	ND
	ant(6)-Ia	1	aminoglycoside	100	Contig_784_46.6268	3050	ND
	aph(3')-III	1	aminoglycoside	99.87	Contig_784_46.6268	3050	ND
	dfrG	1	trimethoprim	89.56	Contig_967_6.80392	chromosome	ND
	msr(C)	1	macrolide	98.99	Contig_623_29.9375	chromosome	ND
	tet(L)	2	tetracycline	99.93	Contig_948_16.7807	chromosome	ND
	VanHAX	2	glycopeptide	99.85	Contig_842_308.46	1649	ND;rep_cluster_889
SEQ-0180	*aac(6')*	1	aminoglycoside	99.64	Contig_336_29.744	chromosome	ND
	ant(6)-Ia	1	aminoglycoside	100	Contig_61_64.4155	3050	ND
	aph(3')-III	1	aminoglycoside	99.87	Contig_61_64.4155	3050	ND
	cat(pC221)	1	phenicol	94.29	Contig_529_42.2446	chromosome	ND
	dfrG	1	trimethoprim	100	Contig_146_19.5467	chromosome	ND
	erm(B)	12	macrolide	100	Contig_506_95.5414	3050	ND
	msr(C)	1	macrolide	98.99	Contig_384_43.616	chromosome	ND
	tet(L)	2	tetracycline	100	Contig_534_20.8178	2351	ND
	tet(M)	10	tetracycline	95.78	Contig_111_46.0329	1866	ND;rep_cluster_185
	VanHAX	2	glycopeptide	99.85	Contig_90_141.675	1649	ND;rep_cluster_889;rep_cluster_893

(continued)

Strain	Gene	Allele	Resistance	% Identity	Contig	Location	Plasmid Incompatibility Sets
SEQ-0181	*aac(6')*	1	aminoglycoside	99.64	Contig_136_112.691	chromosome	ND
	ant(6)-Ia	1	aminoglycoside	100	Contig_385_219.85	1866	ND;rep_cluster_1118;rep_cluster_185
	aph(3')-III	1	aminoglycoside	99.87	Contig_385_219.85	1866	ND;rep_cluster_1118;rep_cluster_185
	cat(pC221)	1	phenicol	97.69	Contig_385_219.85	1866	ND;rep_cluster_1118;rep_cluster_185
	dfrG	1	trimethoprim	100	Contig_224_63.2335	chromosome	ND
	msr(C)	1	macrolide	98.99	Contig_336_149.137	chromosome	ND
	tet(L)	2	tetracycline	100	Contig_447_77.2161	1866	ND;rep_cluster_1118;rep_cluster_185
	tet(M)	10	tetracycline	96.46	Contig_16_192.474	1866	ND;rep_cluster_1118;rep_cluster_185
	VanHAX	2	glycopeptide	99.85	Contig_69_954.984	1649	ND;rep_cluster_889
SEQ-0182	*aac(6')*	1	aminoglycoside	79.96	Contig_993_18.7589	chromosome	ND
	ant(6)-Ia	1	aminoglycoside	100	Contig_1110_73.3041	3050	ND
	aph(3')-III	1	aminoglycoside	99.87	Contig_1110_73.3041	3050	ND
	cat(pC221)	1	phenicol	97.69	Contig_883_101.964	novel_0	rep_cluster_1118
	msr(C)	1	macrolide	98.99	Contig_673_20.7323	chromosome	ND
	tet(L)	2	tetracycline	100	Contig_1313_17.7267	2351	ND
	tet(M)	10	tetracycline	96.46	Contig_863_35.564	1866	ND;rep_cluster_185
	VanHAX	2	glycopeptide	99.85	Contig_232_293.071	1656	ND;rep_cluster_889
SEQ-0184	*aac(6')*	1	aminoglycoside	99.64	Contig_142_16.2675	chromosome	ND
	ant(6)-Ia	1	aminoglycoside	100	Contig_1283_37.2241	3050	ND
	aph(3')-III	1	aminoglycoside	99.87	Contig_1283_37.2241	3050	ND
	cat(pC221)	1	phenicol	88.89	Contig_1254_33.0841	chromosome	ND
	msr(C)	1	macrolide	98.99	Contig_103_16.0565	chromosome	ND
	tet(L)	2	tetracycline	100	Contig_1513_12.5701	2351	ND
	tet(M)	10	tetracycline	95.47	Contig_318_27.4978	1866	ND;rep_cluster_185
	VanHAX	2	glycopeptide	99.85	Contig_633_117.347	1656	ND;rep_cluster_1197;rep_cluster_889;rep_cluster_943

Supplementary Table 10 Virulence factors identified from assembled genomes of AA, B1 and CC VRE isolates

Pulsotype	Isolate	Virulence Factor	Identity	Query/ Template length	Contig	Position in Contig	Accession number
AA	SEQ-0161	*acm*	100	1588 / 2166	Contig_400_15.9256	703..2290	CP003351.1
		efaAfm	90.94	861 / 879	Contig_97_29.6504	6254..7114	AF042288.1
	SEQ-0163	*acm*	100	1444 / 2166	Contig_187_11.0509	1..1444	CP003351.1
		efaAfm	90.94	861 / 879	Contig_667_17.1371	17..877	AF042288.1
	SEQ-0167	*acm*	100	1964 / 2166	Contig_1254_13.8205	1..1964	CP003351.1
		efaAfm	90.94	861 / 879	Contig_916_21.0035	6222..7082	AF042288.1
	SEQ-0169	*acm*	100	1321 / 2166	Contig_5_32.0753	1..1321	CP003351.1
		efaAfm	90.94	861 / 879	Contig_414_68.8532	24671..25531	AF042288.1
	SEQ-0170	*acm*	100	1977 / 2166	Contig_669_22.1201	1..1977	CP003351.1
		efaAfm	90.94	861 / 879	Contig_653_44.8403	24663..25523	AF042288.1
B1	SEQ-0171	*acm*	100	2166 / 2166	Contig_511_45.9406	252..2417	CP003351.1
		efaAfm	90.94	861 / 879	Contig_272_49.2613	13046..13906	AF042288.1
		hylEfm	100	1662 / 1662	Contig_205_59.0231	11873..13534	HM565216.1
	SEQ-0172	*acm*	100	2051 / 2166	Contig_258_85.6327	1..2051	CP003351.1
		efaAfm	90.94	861 / 879	Contig_320_90.127	16697..17557	AF042288.1
		hylEfm	100	1662 / 1662	Contig_119_127.847	11873..13534	HM565216.1
CC	SEQ-0173	*acm*	100	2085 / 2166	Contig_306_49.3024	16..2100	CP003351.1
		efaAfm	90.94	861 / 879	Contig_188_83.783	13455..14315	AF042288.1
		hylEfm	100	1662 / 1662	Contig_177_60.4428	11873..13534	HM565216.1
	SEQ-0174	*acm*	100	2104 / 2166	Contig_232_43.1915	1..2104	CP003351.1
		efaAfm	90.94	861 / 879	Contig_391_68.6669	1715..2575	AF042288.1
		hylEfm	100	1662 / 1662	Contig_167_63.3282	11873..13534	HM565216.1

(continued)

Pulsotype	Isolate	Virulence Factor	Identity	Query/ Template length	Contig	Position in Contig	Accession number
CC	SEQ-0175	*acm*	100	2110 / 2166	Contig_151_49.3155	1..2110	CP003351.1
		efaAfm	90.94	861 / 879	Contig_512_65.2058	24704..25564	AF042288.1
		hylEfm	100	1662 / 1662	Contig_180_66.6324	11873..13534	HM565216.1
	SEQ-0176	*acm*	100	1395 / 2166	Contig_814_13.0014	815..2209	CP003351.1
		efaAfm	90.94	861 / 879	Contig_799_27.3034	28857..29717	AF042288.1
		hylEfm	100	1662 / 1622	Contig_316_25.6938	11873..13534	HM565216.1
	SEQ-0177	*acm*	100	2158 / 2166	Contig_416_19.4457	230..2387	CP003351.1
		efaAfm	90.94	861 / 879	Contig_266_40.9696	24694..25554	AF042288.1
		hylEfm	100	1662 / 1622	Contig_645_35.8309	5660..7321	HM565216.1
	SEQ-0178	*acm*	100	2042 / 2166	Contig_306_42.0905	1..2042	CP003351.1
		efaAfm	90.94	861 / 879	Contig_598_70.7513	24704..25564	AF042288.1
		hylEfm	100	1662 / 1662	Contig_219_92.4115	11873..13534	HM565216.1
	SEQ-0179	*efaAfm*	90.94	861 / 879	Contig_113_31.4402	1688..2548	AF042288.1
		hylEfm	100	1662 / 1662	Contig_316_65.4705	5304..6965	HM565216.1
	SEQ-0182	*acm*	100	2036 / 2166	Contig_903_14.9206	1..2036	CP003351.1
		efaAfm	90.94	861 / 879	Contig_941_17.2419	1043..1903	AF042288.1
		hylEfm	100	1662 / 1662	Contig_969_24.8287	9102..10763	HM565216.1
	SEQ-0184	*acm*	100	2096 / 2166	Contig_1425_9.97832	1..2096	CP003351.1
		efaAfm	90.94	861 / 879	Contig_283_17.7926	1695..2555	AF042288.1
		hylEfm	100	1662 / 1662	Contig_1438_18.0287	766..2427	HM565216.1

Supplementary Table 11 Plasmids identified by PlasmidFinder for VRE isolates using WGS

Isolate	Rep	Plasmid	Identity	Query/ Template length	Contig	Position in contig	Note	Accession number
SEQ-0174	Inc18	rep2	100	1494 / 1494	Contig_387_86.8528	1924..3417	orf1(pRE25)	X92945
SEQ-0175	Inc18	rep2	100	1494 / 1494	Contig_444_134.957	1924..3417	orf1(pRE25)	X92945
SEQ-0176	Inc18	rep2	100	1494 / 1494	Contig_735_108.958	1924..3417	orf1(pRE25)	X92945
SEQ-0177	Inc18	rep2	100	1494 / 1494	Contig_247_70.0747	1810..3303	orf1(pRE25)	X92945
SEQ-0178	Inc18	rep2	100	1494 / 1494	Contig_187_271.645	569..2062	orf1(pRE25)	X92945
SEQ-0179	Inc18	rep2	100	1494 / 1494	Contig_827_137.81	446..1939	orf1(pRE25)	X92945
SEQ-0180	Inc18	rep2	100	1494 / 1494	Contig_298_77.4673	528..2021	orf1(pRE25)	X92945
SEQ-0181	Inc18	rep2	100	1494 / 1494	Contig_126_276.299	596..2089	orf1(pRE25)	X92945
SEQ-0182	Inc18	rep2	100	1494 / 1494	Contig_1183_133.407	1924..3417	orf1(pRE25)	X92945
SEQ-0184	Inc18	rep2	100	1494 / 1494	Contig_901_51.6676	1803..3296	orf1(pRE25)	X92945
SEQ-0162	rep_trans	rep14a	100	768 / 768	Contig_517_8285.66	931..1698	CDS2(pEFNP1)	AB038522
SEQ-0164	rep_trans	rep14a	100	768 / 768	Contig_150_8147.43	931..1698	CDS2(pEFNP1)	AB038522
SEQ-0167	rep_trans	rep14a	100	768 / 768	Contig_495_5980.6	931..1698	CDS2(pEFNP1)	AB038522
SEQ-0169	rep_trans	rep14a	100	768 / 768	Contig_176_35346.5	931..1698	CDS2(pEFNP1)	AB038522
SEQ-0170	rep_trans	rep14a	100	768 / 768	Contig_290_24510.7	931..1698	CDS2(pEFNP1)	AB038522
SEQ-0171	rep_trans	rep14a	100	768 / 768	Contig_493_7510.21	614..1381	CDS2(pEFNP1)	AB038522
SEQ-0184	rep_trans	rep14a	100	768 / 768	Contig_590_11487	931..1698	CDS2(pEFNP1)	AB038522
SEQ-0174	rep_trans	rep14a	100	871 / 897	Contig_379_277.65	342..1212	EFAU085p500 (AUS0085p5)	CP006625
SEQ-0175	rep_trans	rep14a	100	871 / 897	Contig_481_808.561	343..1213	EFAU085p5001 (AUS0085p5)	CP006625
SEQ-0176	rep_trans	rep14a	100	871 / 897	Contig_28_397.624	279..1149	EFAU085p5001 (AUS0085p5)	CP006625
SEQ-0177	rep_trans	rep14a	100	871 / 897	Contig_135_195.976	245..1115	EFAU085p5001 (AUS0085p5)	CP006625
SEQ-0178	rep_trans	rep14a	100	871 / 897	Contig_614_1585.65	354..1224	EFAU085p5001 (AUS0085p5)	CP006625
SEQ-0179	rep_trans	rep14a	100	871 / 897	Contig_366_523.13	307..1177	EFAU085p5001 (AUS0085p5)	CP006625

(continued)

Isolate	Rep	Plasmid	Identity	Query/ Template length	Contig	Position in contig	Note	Accession number
SEQ-0180	rep_trans	rep14a	100	871 / 897	Contig_6_253.955	281..1151	EFAU085p5001 (AUS0085p5)	CP006625
SEQ-0181	rep_trans	rep14a	100	871 / 897	Contig_265_1407.89	348..1218	EFAU085p5001 (AUS0085p5)	CP006625
SEQ-0182	rep_trans	rep14a	100	555 / 897	Contig_1215_205.321	1..555	EFAU085p5001 (AUS0085p5)	CP006625
SEQ-0161	rep_trans	rep14a	99.5	402 / 402	Contig_5_11251_Circ	728..1129	ORF1(pKQ10)	EFU01917
SEQ-0162	rep_trans	rep14a	99.5	402 / 402	Contig_3_8075.1_Circ	728..1129	ORF1(pKQ10)	EFU01917
SEQ-0164	rep_trans	rep14a	99.5	402 / 402	Contig_2_8485.46_Circ	728..1129	ORF1(pKQ10)	EFU01917
SEQ-0167	rep_trans	rep14a	99.5	402 / 402	Contig_4_6214.76_Circ	728..1129	ORF1(pKQ10)	EFU01917
SEQ-0169	rep_trans	rep14a	99.5	402 / 402	Contig_4_34416.3_Circ	728..1129	ORF1(pKQ10)	EFU01917
SEQ-0170	rep_trans	rep14a	99.5	402 / 402	Contig_6_24340.8_Circ	728..1129	ORF1(pKQ10)	EFU01917
SEQ-0184	rep_trans	rep14a	99.5	402 / 402	Contig_5_10767.2_Circ	728..1129	ORF1(pKQ10)	EFU01917
SEQ-0171	rep_trans	rep14b	100	966 / 966	Contig_356_8127.15	440..1405	EFAU085p6001 (AUS0085p6)	CP006626
SEQ-0172	rep_trans	rep14b	100	966 / 966	Contig_4_17627_Circ	987..1952	EFAU085p6001 (AUS0085p6)	CP006626
SEQ-0174	rep_trans	rep7a	100	939 / 939	Contig_116_58.0949	628..1566	ORF11(pRE25)	X92945
SEQ-0175	rep_trans	rep7a	100	939 / 939	Contig_318_149.558	489..1427	ORF11(pRE25)	X92945
SEQ-0176	rep_trans	rep7a	100	939 / 939	Contig_535_96.6345	489..1427	ORF11(pRE25)	X92945
SEQ-0177	rep_trans	rep7a	100	939 / 939	Contig_201_26.1158	122..1060	ORF11(pRE25)	X92945
SEQ-0178	rep_trans	rep7a	100	939 / 939	Contig_382_239.889	489..1427	ORF11(pRE25)	X92945
SEQ-0179	rep_trans	rep7a	100	939 / 939	Contig_127_129.346	145..1083	ORF11(pRE25)	X92945
SEQ-0180	rep_trans	rep7a	100	939 / 939	Contig_333_68.5834	489..1427	ORF11(pRE25)	X92945
SEQ-0181	rep_trans	rep7a	100	939 / 939	Contig_385_219.85	1576..2514	ORF11(pRE25)	X92945
SEQ-0182	rep_trans	rep7a	100	939 / 939	Contig_883_101.964	489..1427	ORF11(pRE25)	X92945
SEQ-0184	rep_trans	rep7a	100	939 / 939	Contig_999_81.1592	489..1427	ORF11(pRE25)	X92945
SEQ-0161	rep_trans	repUS43	100	1206 / 1206	Contig_794_43.8738	2437..3642	CDS12738(DOp1)	CP003584
SEQ-0162	rep_trans	repUS43	100	1206 / 1206	Contig_111_30.5274	2581..3786	CDS12738(DOp1)	CP003584
SEQ-0164	rep_trans	repUS43	100	1206 / 1206	Contig_108_86.6925	63..1268	CDS12738(DOp1)	CP003584

(continued)

Isolate	Rep	Plasmid	Identity	Query/ Template length	Contig	Position in contig	Note	Accession number
SEQ-0167	rep_trans	repUS43	100	1206 / 1206	Contig_411_21.168	2550..3755	CDS12738(DOp1)	CP003584
SEQ-0169	rep_trans	repUS43	100	1206 / 1206	Contig_231_93.0658	2461..3666	CDS12738(DOp1)	CP003584
SEQ-0170	rep_trans	repUS43	100	1206 / 1206	Contig_339_53.2034	2430..3635	CDS12738(DOp1)	CP003584
SEQ-0171	rep_trans	repUS43	100	1206 / 1206	Contig_383_51.849	10681..11886	CDS12738(DOp1)	CP003584
SEQ-0172	rep_trans	repUS43	100	1206 / 1206	Contig_223_106.398	12169..13374	CDS12738(DOp1)	CP003584
SEQ-0174	rep_trans	repUS43	100	1206 / 1206	Contig_57_65.8848	2815..4020	CDS12738(DOp1)	CP003584
SEQ-0175	rep_trans	repUS43	100	1206 / 1206	Contig_354_90.6888	2816..4021	CDS12738(DOp1)	CP003584
SEQ-0176	rep_trans	repUS43	100	1206 / 1206	Contig_96_27.7678	2815..4020	CDS12738(DOp1)	CP003584
SEQ-0177	rep_trans	repUS43	100	1206 / 1206	Contig_278_50.1813	2476..3681	CDS12738(DOp1)	CP003584
SEQ-0178	rep_trans	repUS43	100	1206 / 1206	Contig_79_65.3966	2815..4020	CDS12738(DOp1)	CP003584
SEQ-0179	rep_trans	repUS43	100	1206 / 1206	Contig_487_37.3985	2540..3745	CDS12738(DOp1)	CP003584
SEQ-0180	rep_trans	repUS43	100	1206 / 1206	Contig_63_37.9183	2815..4020	CDS12738(DOp1)	CP003584
SEQ-0181	rep_trans	repUS43	100	1206 / 1206	Contig_325_148.374	2816..4021	CDS12738(DOp1)	CP003584
SEQ-0182	rep_trans	repUS43	100	1206 / 1206	Contig_527_25.6599	2637..3842	CDS12738(DOp1)	CP003584
SEQ-0184	rep_trans	repUS43	100	1206 / 1206	Contig_992_22.8785	2447..3652	CDS12738(DOp1)	CP003584
SEQ-0161	rep1	repUS12	99.62	792 / 795	Contig_663_12.668	3413..4204	rep(pUB110)	AF181950
SEQ-0164	rep1	repUS12	99.62	792 / 795	Contig_98_55.4671	3458..4249	rep(pUB110)	AF181950
SEQ-0167	rep1	repUS12	99.62	792 / 795	Contig_961_13.004	3416..4207	rep(pUB110)	AF181950
SEQ-0169	rep1	repUS12	99.62	792 / 795	Contig_295_36.786	3932..4723	rep(pUB110)	AF181950
SEQ-0171	rep1	repUS12	99.62	792 / 795	Contig_197_27.9486	4286..5077	rep(pUB110)	AF181950
SEQ-0162	rep1	repUS12	99.84	610 / 876	Contig_1002_7.50162	29..638	repB(SAP014A)	GQ900379
SEQ-0174	rep1	repUS12	100	531 / 876	Contig_335_22.6174	1011..1541	repB(SAP014A)	GQ900379
SEQ-0175	rep1	repUS12	100	531 / 876	Contig_383_28.2433	1011..1541	repB(SAP014A)	GQ900379
SEQ-0176	rep1	repUS12	100	531 / 876	Contig_640_13.7837	1011..1541	repB(SAP014A)	GQ900379
SEQ-0177	rep1	repUS12	100	534 / 876	Contig_335_37.3832	2302..2835	repB(SAP014A)	GQ900379
SEQ-0178	rep1	repUS12	100	531 / 876	Contig_469_39.6673	1011..1541	repB(SAP014A)	GQ900379
SEQ-0180	rep1	repUS12	100	531 / 876	Contig_409_21.3853	1011..1541	repB(SAP014A)	GQ900379
SEQ-0181	rep1	repUS12	100	531 / 876	Contig_359_66.069	1011..1541	repB(SAP014A)	GQ900379

(continued)

Isolate	Rep	Plasmid	Identity	Query/ Template length	Contig	Position in contig	Note	Accession number
SEQ-0161	rep3	rep11a	100	945 / 945	Contig_294_890.191	9109..10053	repA(pB82)	AB178871
SEQ-0162	rep3	rep11a	100	945 / 945	Contig_543_789.022	9109..10053	repA(pB82)	AB178871
SEQ-0164	rep3	rep11a	100	945 / 945	Contig_156_1580.03	9109..10053	repA(pB82)	AB178871
SEQ-0167	rep3	rep11a	100	945 / 945	Contig_529_856.385	9109..10053	repA(pB82)	AB178871
SEQ-0169	rep3	rep11a	100	945 / 945	Contig_187_2902.66	9109..10053	repA(pB82)	AB178871
SEQ-0170	rep3	rep11a	100	945 / 945	Contig_301_2261.3	9109..10053	repA(pB82)	AB178871
SEQ-0174	rep3	rep11a	100	857 / 945	Contig_3_1158.08_Circ	5317..6173	repA(pB82)	AB178871
SEQ-0175	rep3	rep11a	100	945 / 945	Contig_221_3463.4	1979..2923	repA(pB82)	AB178871
SEQ-0176	rep3	rep11a	100	945 / 945	Contig_386_2293.68	1979..2923	repA(pB82)	AB178871
SEQ-0177	rep3	rep11a	100	857 / 945	Contig_4_1920.19_Circ	5317..6173	repA(pB82)	AB178871
SEQ-0178	rep3	rep11a	100	945 / 945	Contig_274_7206.49	1979..2923	repA(pB82)	AB178871
SEQ-0179	rep3	rep11a	100	857 / 945	Contig_5_4567.34_Circ	5317..6173	repA(pB82)	AB178871
SEQ-0180	rep3	rep11a	100	945 / 945	Contig_217_1432.44	1979..2923	repA(pB82)	AB178871
SEQ-0181	rep3	rep11a	100	945 / 945	Contig_196_5961.82	1979..2923	repA(pB82)	AB178871
SEQ-0182	rep3	rep11a	100	945 / 945	Contig_565_2597.49	1979..2923	repA(pB82)	AB178871
SEQ-0183	rep3	rep11a	100	945 / 945	Contig_1511_846.128	741..1685	repA(pB82)	AB178871
SEQ-0184	rep3	rep11a	100	945 / 945	Contig_835_1532.78	1901..2845	repA(pB82)	AB178871
SEQ-0174	rep3	rep18b	100	969 / 969	Contig_310_803.438	4594..5562	BO23315710(pE1p13)	CP018068
SEQ-0175	rep3	rep18b	100	969 / 969	Contig_2_2505.07_Circ	2363..3331	BO23315710(pE1p13)	CP018068
SEQ-0176	rep3	rep18b	100	969 / 969	Contig_1_1799.76_Circ	2363..3331	BO23315710(pE1p13)	CP018068
SEQ-0177	rep3	rep18b	100	969 / 969	Contig_2_1073.91_Circ	2363..3331	BO23315710(pE1p13)	CP018068
SEQ-0178	rep3	rep18b	100	969 / 969	Contig_2_5493.92_Circ	2363..3331	BO23315710(pE1p13)	CP018068
SEQ-0179	rep3	rep18b	100	969 / 969	Contig_3_2728.08_Circ	2363..3331	BO23315710(pE1p13)	CP018068
SEQ-0180	rep3	rep18b	100	969 / 969	Contig_2_1044.55_Circ	2363..3331	BO23315710(pE1p13)	CP018068
SEQ-0181	rep3	rep18b	100	969 / 969	Contig_2_4450.39_Circ	2363..3331	BO23315710(pE1p13)	CP018068
SEQ-0182	rep3	rep18b	100	969 / 969	Contig_2_2058.43_Circ	2363..3331	BO23315710(pE1p13)	CP018068
SEQ-0183	rep3	rep18b	100	969 / 969	Contig_4_586.881	5120..6088	BO23315710(pE1p13)	CP018068
SEQ-0184	rep3	rep18b	100	969 / 969	Contig_89_1208.55	5350..6318	BO23315710(pE1p13)	CP018068

(continued)

Isolate	Rep	Plasmid	Identity	Query/ Template length	Contig	Position in contig	Note	Accession number
SEQ-0161	rep3	rep18b	99.43	527 / 564	Contig_294_890.191	960..1486	repA(pEF418)	AF408195
SEQ-0162	rep3	rep18b	99.43	527 / 564	Contig_543_789.022	960..1486	repA(pEF418)	AF408195
SEQ-0164	rep3	rep18b	99.43	527 / 564	Contig_156_1580.03	960..1486	repA(pEF418)	AF408195
SEQ-0165	rep3	rep18b	99.43	527 / 564	Contig_2491_272.543	872..1398	repA(pEF418)	AF408195
SEQ-0166	rep3	rep18b	99.43	527 / 564	Contig_1951_174.817	872..1398	repA(pEF418)	AF408195
SEQ-0167	rep3	rep18b	99.43	527 / 564	Contig_529_856.385	960..1486	repA(pEF418)	AF408195
SEQ-0169	rep3	rep18b	99.43	527 / 564	Contig_187_2902.66	960..1486	repA(pEF418)	AF408195
SEQ-0170	rep3	rep18b	99.43	527 / 564	Contig_301_2261.3	960..1486	repA(pEF418)	AF408195
SEQ-0171	rep3	rep18b	99.43	527 / 564	Contig_2_2157.45_Circ	2399..2925	repA(pEF418)	AF408195
SEQ-0172	rep3	rep18b	99.43	527 / 564	Contig_2_4772.5_Circ	2399..2925	repA(pEF418)	AF408195
SEQ-0162	repA_N	rep17	100	1041 / 1041	Contig_860_35.7661	555..1595	CDS29(pRUM)	AF507977
SEQ-0164	repA_N	rep17	100	989 / 1041	Contig_194_119.467	1..989	CDS29(pRUM)	AF507977
SEQ-0165	repA_N	rep17	100	698 / 1041	Contig_911_16.6681	1..698	CDS29(pRUM)	AF507977
SEQ-0167	repA_N	rep17	100	981 / 1041	Contig_926_42.6844	1..981	CDS29(pRUM)	AF507977
SEQ-0169	repA_N	rep17	100	1041 / 1041	Contig_276_119.167	555..1595	CDS29(pRUM)	AF507977
SEQ-0170	repA_N	rep17	100	1015 / 1041	Contig_97_56.3339	186..1200	CDS29(pRUM)	AF507977
SEQ-0171	repA_N	rep17	100	1041 / 1041	Contig_116_31.6862	310..1350	CDS29(pRUM)	AF507977
SEQ-0172	repA_N	rep17	100	1041 / 1041	Contig_144_73.4716	560..1600	CDS29(pRUM)	AF507977
SEQ-0174	repA_N	rep17	100	1041 / 1041	Contig_203_58.9854	560..1600	CDS29(pRUM)	AF507977
SEQ-0175	repA_N	rep17	100	1041 / 1041	Contig_74_140.162	6873..7913	CDS29(pRUM)	AF507977
SEQ-0176	repA_N	rep17	100	1041 / 1041	Contig_624_36.1667	21..1061	CDS29(pRUM)	AF507977
SEQ-0177	repA_N	rep17	100	1041 / 1041	Contig_495_31.1539	294..1334	CDS29(pRUM)	AF507977
SEQ-0178	repA_N	rep17	100	1041 / 1041	Contig_280_100.539	560..1600	CDS29(pRUM)	AF507977
SEQ-0179	repA_N	rep17	100	1041 / 1041	Contig_601_39.5364	178..1218	CDS29(pRUM)	AF507977
SEQ-0180	repA_N	rep17	100	1041 / 1041	Contig_223_71.3212	560..1600	CDS29(pRUM)	AF507977
SEQ-0181	repA_N	rep17	100	1041 / 1041	Contig_202_239.704	560..1600	CDS29(pRUM)	AF507977
SEQ-0182	repA_N	rep17	100	1041 / 1041	Contig_577_77.5506	560..1600	CDS29(pRUM)	AF507977
SEQ-0183	repA_N	rep17	100	693 / 1041	Contig_2218_18.214	1..693	CDS29(pRUM)	AF507977

(continued)

Isolate	Rep	Plasmid	Identity	Query/ Template length	Contig	Position in contig	Note	Accession number
SEQ-0184	repA_N	rep17	99.9	1041 / 1401	Contig_174_43.8762	6873..7912	CDS29(pRUM)	AF507977
SEQ-0162	repA_N	repUS15	100	656 / 1041	Contig_1038_15.3429	1..656	repA(pNB2354p1)	CP004064
SEQ-0164	repA_N	repUS15	100	1041 / 1041	Contig_324_44.0575	419..1459	repA(pNB2354p1)	CP004064
SEQ-0167	repA_N	repUS15	100	1041 / 1041	Contig_1286_18.7289	67..1107	repA(pNB2354p1)	CP004064
SEQ-0171	repA_N	repUS15	100	1041 / 1041	Contig_289_28.5676	129..1169	repA(pNB2354p1)	CP004064
SEQ-0172	repA_N	repUS15	100	1041 / 1041	Contig_80_52.1678	325..1365	repA(pNB2354p1)	CP004064
SEQ-0175	repA_N	repUS15	99.9	1041 / 1041	Contig_374_41.0507	118..1158	repA(pNB2354p1)	CP004064
SEQ-0176	repA_N	repUS15	99.9	1041 / 1041	Contig_775_19.0616	131..1171	repA(pNB2354p1)	CP004064
SEQ-0180	repA_N	repUS15	99.9	1041 / 1041	Contig_91_23.6152	343..1383	repA(pNB2354p1)	CP004064
SEQ-0182	repA_N	repUS15	99.9	1041 / 1041	Contig_183_14.5922	2..1042	repA(pNB2354p1)	CP004064

Supplementary Table 12 Tabulation of ^{1}H NMR spectral band assignments reported in the literature

^{1}H δppm	**Possible biomolecule contributor**	**Reference**
0.72s	Angular methyl groups in cholesterol	[1]
0.8	Methyl- (-CH_3) lipid – distinctive signals emitted by terminal methyl group protons of phospholipids, chloesterol, cholesterol esters, TGs (from lipoproteins)	[2]
0.88	Fatty acid CH_3	[3, 4]
0.89	Phospholipids CH_3	[5]
0.89-0.96	α-hydroxybutyrate, valine, leucine m CH_3 isoleucine t CH_3	[3, 6, 7]
0.93	Methyl (-CH_3) all fatty acids Isoleucine t CH_3	[8] [9]
0.99d-1.02	Isoleucine CH_3	[3, 9]
1.00-1.06d	Valine CH_3	[3, 7]
1.02s	Angular methyl groups in cholesterol	[1]
1.19t	ethanol	[3, 10]
1.2	Methylene (-CH_2-) lipid	[2]
1.27	Phospholipid CH_2	[5]
1.32	-$(CH_2)_n$- all fatty acids (except unsatureated FA)	[8]
1.31-1.34d	Lactate CH_3	[3, 6, 11]
1.34d	threonine	[10]
1.47-1.48d	Alanine CH	[9, 11]
1.61	–C*H*2–CH2–COOR all fatty acids (except unsaturated FA) Citrulline m	[8] [10]
1.71m	Arginine CH_2 Lysine CH_2	[9]
1.80s	acetate	[7]
1.87m	Citrulline CH_2	[9]
1.89-1.92	Acetate, glutamine Arginine (m, CH_2), lysine (mCH_2)	[6, 11] [9]
1.97	Isoleucine Proline (m) CH_2	[9]
2.00	Lipid CH=CHCH_2CH=CH	[4]
2.01s	N-Acetyl groups in polysaccharides	[5]
2.05	–C*H2*–CH=CH– unsaturated FA	[8]
2.06	Proline (m) CH_2 Glutamate	[9]

(continued)

^{1}H δppm	Possible biomolecule contributor	Reference
2.08	Ketoleucine, N-Acetyl groups in polysaccharides	[5]
2.11m	Glutamine CH_2	[9]
2.12m	Glutamate CH_2	[9]
2.26	Valine	
2.25-2.27	glutamine	[6]
2.25-2.31	glutamine	[6]
2.27	–CH2–COOR all FA	[8]
2.30	GABA (γ-aminobutyric acid)	[10]
2.32m	Glutamate CH_2	[9]
2.34m	Proline CH_2	[9]
2.35	Lipid CH_2COO	[4]
2.39	Pyruvate	[11]
2.37dd	Malic acid α-CH_2	[9]
2.40s	Succinic acid CH_2	[9, 11]
2.43m	Glutamine CH_2	[9]
2.53d	Citrate α-CH_2	[9]
2.66-2.70	Citrate doublet α'-CH_2 Malic acid dd α'-CH_2	[6, 9, 10]
2.86	=CH–C*H*2–CH= polyunsaturated FA Asparagine (m, CH)	[8] [9]
2.95m	Asparagine CH	[9]
3.04s	Creatine, phosphocreatine N-CH_3	[3]
3.11	Carnosine CH_2	[3]
3.13	phosphorylcholine	[7]
3.14-3.15	Citrulline, Tyrosine	[9]
3.16m	PE	[1]
3.18s	Choline **N(CH_3)**	[3, 9]
3.20	PI (t), PE (bs)	[1]
3.2-4.34	Inositol of PI	[12]
3.22s	PC	[1]
3.21-3.25*	Choline/phosphatidylcholine/O-phosphocholine ($N^+(CH_3)_3$)	[3, 5, 9, 11]
3.23	Arginine CH_2 (t) Glycinebetaine (s)	[9] [11]
3.23-3.25	Glucose dd CH	[6, 9]
3.24s	phosphorylcholine	[13]
3.24-3.25s	Glycerophosphocholine phosphoethanoalmine	[13, 14]
3.25-3.27	Betaine	[5, 7]

(continued)

^{1}H δppm	Possible biomolecule contributor	Reference
3.27s	Phosphatidylcholine Phosphlipids $N(CH_3)_3$	[5, 14]
3.30s	PC	[1]
3.33dt	Proline CH_2	[9]
3.48t	Glucose CH	[9]
3.51dd	Choline CH_2	[9]
3.52dd	Glucose CH Myo-inositol CH	[9]
3.52-3.83	Glycerol of CL	[12]
3.53d	leucine	[7]
3.56	Arabinitol -H1, H5	[15]
3.60	Phosphocholine NCH_2	[5]
3.62-3.72	Arabinitol -H3, Mannitol-H3, H4 Myo-inositol	[15] [13]
3.62	phosphocholine	[14]
3.65-3.68m	glycerophosphocholine	[13, 14]
3.68	α glycerophosphate	[16]
3.69	Ethanolamine phosphate	[16]
3.70t	Glucose CH	[9]
3.74	Citrulline dd CH Glutamine dd CH_2 Lysine t CH Malic acid t CH	[9]
3.75	α glycerophosphate	[16]
3.76t	Arginine CH	[9]
3.77q	Alanine CH_3	[9]
3.79	phosphatidylcholine	[16]
3.79-3.91	Central glycerol of PG $POCH_2CH(OH)CH_2OP$	[12]
3.81	phosphatidylethanolamine	[16]
3.84	Guanosine 5’ monophosphate	[16]
3.88	$CH_2N^+(CH_3)_3$ in PC, LPC diphosphate	[1, 16]
3.89-3.90	Fructose dd CH Betaine s Guanosine 5’ monophosphate Adenosine 5’ monophosphate	[9] [5, 7]
3.91-3.93	CH3 in glycerol Polyols, carbohydrate residues C_3H_2OP glycerol backbone	[1, 6, 12]

(continued)

^{1}H δppm	Possible biomolecule contributor	Reference
3.93	Arabinitol -H2, H4 Creatine/glycerophosphocholine	[15] [13, 14]
3.94	Backbone of PC, diphosphate	[1, 16]
3.96m	Glycerol of phospholipids	[1]
3.97	phosphoethanolamine	[14]
4.00dd	Asparagine CH	[9]
4.02t	Fructose CH_2	[9]
4.05	Glycerol of phospholipids	[1]
4.07	β glycerophosphate	[16]
4.08	Choline phosphate	[16]
4.10ddd	Choline CH_2	[9]
4.11	–CH2–OCOR glyceryl backbone	[8, 12]
4.17-4.19t	Phosphocholine α-CH_2	[3, 5]
4.23	phosphatidylcholine	[16]
4.25	DAG CH_2OCO	[4]
4.28	TAG CH_2OCO	[4]
4.31	–CH2–OCOR′ glyceryl	[8]
4.33-4.34t	Glycerophosphorylcholine α-CH_3 POC_4H inosphitol of PI	[3] [12]
4.35	Lipid CH_2COO glycerol backbone	[12]
4.38	Uridine 3’ monophosphate	[16]
4.43q	Plasmenyl PC and/or plasmenyl PE	[1]
4.51	ATR -CH Rib	[17]
4.52s	myoinositol	[16]
4.54	Uridine 2’ monophosphate	[16]
4.58	Guanosine 3’ monophosphate	[16]
4.64d	Glucose CH Maltose CH	[9]
4.8	ATP -CH Rib	[17]
4.89	Guanosine 2’ monophosphate	[16]
4.93	Adenosine 2’ 5’	[16]
5.14	C_2HOC lipid glycerol backbones	[12]
5.17d	α-glucose	[7]
5.22d	Glucose CH Maltose CH	[9]
5.33-5.38	–CH=CH– unsaturated FA	[3, 4, 8]
5.34	α-D-glucose-1-phosphate	[16]
5.40d	Maltose CH	[9]

(continued)

^{1}H δppm	Possible biomolecule contributor	Reference
5.45d	sphingomyelin	[1]
~5.5m	UGLcNAc, UGalNAc, UGluA	[18]
	Ribose-5-phosphate	[7]
5.57	ergosterol	[4]
5.69d	uridine	[7]
5.71dt	sphingomylelin	[1]
5.79d	uridine	[7]
5.96d	Plasmenyl PC and/or plasmenyl PE	[1]
6.15d	ADP/ATP -CH Rib	[11, 17]
6.50s	Fumarate CH=CH	[9]
6.52	Fumaric acid	[10]
7.42d	uridine	[7]
8.18s	NAD	[11]
8.20t	NAD	[11]
8.22	GMP	[4]
8.25s	AMP CH-8	[3]
8.26	ATP CH	[17]
8.35s	Adenosine	[11]
8.46s	formate	[11]
8.49s	AMP CH-2	[3]
8.54s	ATP -CH	[17]
8.81-8.84d	NAD -CH Nic	[11, 17]
9.12-9.15	NAD -CH Nic	[11, 17]
9.33-9.34	NAD -CH Nic	[11, 17]

References in Appendix

1. Li, J., T. Vosegaard, and Z. Guo, *Applications of nuclear magnetic resonance in lipid analyses: An emerging powerful tool for lipidomics studies.* Progress in lipid research, 2017. **68**: p. 37-56.
2. Aru, V., et al., *Quantification of lipoprotein profiles by nuclear magnetic resonance spectroscopy and multivariate data analysis.* TrAC Trends in Analytical Chemistry, 2017. **94**: p. 210-219.
3. García-García, A.B., et al., *1H HR-MAS NMR-based metabolomics analysis for dry-fermented sausage characterization.* Food chemistry, 2018. **240**: p. 514-523.
4. Palomino-Schätzlein, M., et al., *Optimised protocols for the metabolic profiling of S. cerevisiae by 1 H-NMR and HRMAS spectroscopy.* Analytical and bioanalytical chemistry, 2013. **405**(26): p. 8431-8441.
5. Righi, V., et al., *Effects of a small, volatile bacterial molecule on Pseudomonas aeruginosa bacteria using whole cell high-resolution magic angle spinning nuclear magnetic resonance spectroscopy and genomics.* International journal of molecular medicine, 2018. **42**(4): p. 2129-2136.
6. Himmelreich, U., et al., *Rapid etiological classification of meningitis by NMR spectroscopy based on metabolite profiles and host response.* PloS one, 2009. **4**(4): p. e5328.
7. Liu, Q., et al., *Evaluation of the metabolic response of Escherichia coli to electrolysed water by 1H NMR spectroscopy.* LWT-Food Science and Technology, 2017. **79**: p. 428-436.
8. Merkley, N. and R.T. Syvitski, *Profiling whole microalgal cells by high-resolution magic angle spinning (HR-MAS) magnetic resonance spectroscopy.* Journal of applied phycology, 2012. **24**(3): p. 535-540.
9. Pagter, M., C.C. Yde, and K.H. Kjær, *Metabolic fingerprinting of dormant and active flower primordia of ribes nigrum using high-resolution magic angle spinning NMR.* Journal of agricultural and food chemistry, 2017. **65**(46): p. 10123-10130.
10. da Conceição Santos, A.D., et al., *1H HR-MAS NMR-based metabolomics study of different persimmon cultivars (Diospyros kaki) during fruit development.* Food chemistry, 2018. **239**: p. 511-519.
11. Bundy, J.G., et al., *Discrimination of pathogenic clinical isolates and laboratory strains of Bacillus cereus by NMR-based metabolomic profiling.* FEMS microbiology letters, 2005. **242**(1): p. 127-136.
12. Mastronicolis, S., et al., *Coordinated regulation of cold-induced changes in fatty acids with cardiolipin and phosphatidylglycerol composition among phospholipid species for the food pathogen Listeria monocytogenes.* Applied and environmental microbiology, 2008. **74**(14): p. 4543-4549.
13. Chihanga, T., et al., *Influence of media selection on NMR based metabolic profiling of human cell lines.* Metabolomics, 2018. **14**(3): p. 28.
14. Duarte, I.F., et al., *Analytical approaches toward successful human cell metabolome studies by NMR spectroscopy.* Analytical chemistry, 2009. **81**(12): p. 5023-5032.
15. Alcantara, G.B., et al., *Chemometric analysis applied in 1H HR-MAS NMR and FT-IR data for chemotaxonomic distinction of intact lichen samples.* Analytica chimica acta, 2007. **595**(1-2): p. 3-8.

16. Vestergren, J., et al., *High-resolution characterization of organic phosphorus in soil extracts using 2D 1H–31P NMR correlation spectroscopy.* Environmental Science & Technology, 2012. **46**(7): p. 3950-3956.
17. Vermathen, M., et al., *1H HR-MAS NMR spectroscopy to study the metabolome of the protozoan parasite Giardia lamblia.* Talanta, 2018. **188**: p. 429-441.
18. Vermathen, M., et al., *Metabolic Profiling of Cells in Response to Drug Treatment using 1H High-resolution Magic Angle Spinning (HR-MAS) NMR Spectroscopy.* CHIMIA International Journal for Chemistry, 2017. **71**(3): p. 124-129.